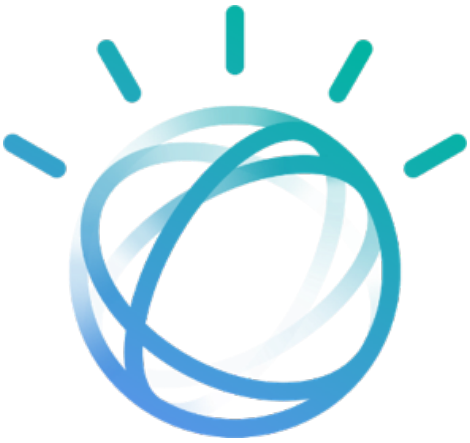
h e g

Haute école de gestion
Genève

Architecture des applications utilisant Watson.

Cas pratique de l'interrogation d'une base de documents



Travail de bachelor réalisé en vue de l'obtention du Bachelor HEG

par :

Bassim BENTAHAR

Directeur de mémoire:

Philippe Dugerdil, Professeur HES

Lausanne, le 01 septembre 2017

Haute école de gestion de Genève (HEG-GE)

Filière informatique de gestion

Hes·SO GENÈVE

DECLARATION

Ce travail de bachelor est réalisé dans le cadre de l'examen final de la haute école de gestion de Genève, en vue de l'obtention du titre « Bachelor of Science en informatique de Gestion ».

L'étudiante atteste que son travail a été vérifié par un logiciel de détection de plagiat.

L'étudiante accepte, le cas échéant, la clause de confidentialité. L'utilisation des conclusion et recommandations formulées dans le travail de Bachelor, sans préjuger de leur valeur, n'engage ni la responsabilité de l'auteur ni celle de conseiller au travail de Bachelor, du juré et de la HEG.

" J'atteste avoir réalisé seul le présent travail, sans avoir utilisé des sources autre que celle citées dans la bibliographie."

Fait à Lausanne, le 01 septembre 2017

Bassim Bentahar

REMERCIEMENTS

Je souhaite remercier la haute école de Gestion ainsi les professeurs m'ayant soutenu durant c'est 4 années de formation.

J'aimerais particulièrement remercier mon directeur de mémoire, Monsieur Philippe Dugerdil qui m'a beaucoup aidé concernent les pistes de réflexion à suivre. Ses conseils, son engagement, son enthousiasme, son expérience et ses connaissances mon accompagné tout au long de ce travail.

Je remercie profondément mes parents, et mes frères, pour leur soutient inconditionnel, je remercie également ma femme qui a su me soutenir non seulement pendant ce travail, mais aussi pendant les 4 ans de formation.

RESUME

L'idée initiale de ce Travail de Bachelor était de créer une application vocale qui utilise les APIs cognitifs de Watson pour interroger une base de documents. Il s'est avéré, après une étude préliminaire, que la création d'une application qui utilise ces APIs cognitifs n'était que la dernière étape dans un grand parcours qui commence par connaître les spécificités des recherches classiques puis le fonctionnement d'un champ de l'intelligence artificielle, à savoir l'apprentissage d'ordonnancement (*Learning to rank*) et, enfin, maîtriser l'entraînement du système pour répondre à la spécificité de l'application visée. En fait, l'entraînement est le point essentiel, car la façon d'entrainer le système — qualitatif et quantitatif — influence fortement la pertinence des documents retournés.

La partie vocale a été laissé de côté pour deux raisons :

- Pour ne pas disperser le lecteur et se concentrer sur l'étude des recherches textuelles.
- Le service de transformation de voix en texte a pas mal de lacunes pour être intégré avec une recherche textuelle.

Pour cette raison nous avons tenté à travers ce document de présenter l'essentiel des connaissances préalables aux développements d'une application cognitive de recherche textuelle.

Le travail peut être coupé essentiellement en trois axes :

- Les éléments essentiels de Solr qui implémente la recherche classique de la partie Retrieve,
- Ce qu'il faut savoir pour entraîner le système pour la partie Rank,
- Test de performance du système à travers un jeu de tests pour ressortir quelques conclusions.

TABLE DES MATIERES

LISTE DES FIGURES

I. INTRODUCTION

Le terme de l'intelligence artificielle symbolise la grande espérance dans le monde informatique du moment. C'est que, tout semble destiné à être révolutionné par cette technologie qui suscite autant d'espoirs que de craintes. Depuis quelques années, les ordinateurs font des progrès dans la maitrise des jeux comme les échecs[1], le jeu de go[2] , le jeu télévisé Jeopardy![3] ainsi que la reconnaissance d'image. Qu'en est-il dans la recherche d'information textuelle ?

Actuellement trois états de fait encouragent les data scientists à se pencher sur les recherches textuelles utilisant le « Machin learning »[4] pour la recherche de l'information pertinente vis-à-vis d'une requête :

- La disponibilité d'immenses corpus numérisés sous format non structurés, ces données se multipliant[5] par 100 chaque 10 ans. Par données non structurées, nous entendons toute donnée extérieure à un type de structure.

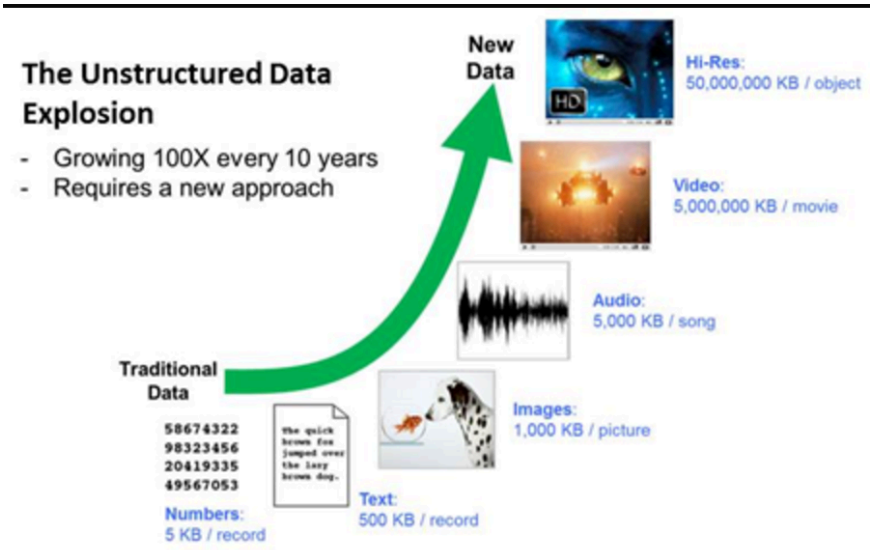


FIGURE 1: EXPLOSION DES DONNEES NON STRUCTUREES

[1] Deep blue a battu le champion du monde Garry kasparov en 1997

[2] AlphaGo a battu le champion du monde dans le jeu de go Ke Jie

[3] Watson a battu les deux plus grands champions dans le jeu télévisé Joepardy !

[4] https://fr.wikipedia.org/wiki/Apprentissage_automatique

[5] http://www.imexresearch.com/newsletters/obs.html

- Les recherches par mots clés sur les données non structurées présentent beaucoup de lacunes car elles ne prennent pas en compte plusieurs aspects qui peuvent s'avérer très importants lors des recherches d'informations.
- Selon la Loi de Moor : « La puissance des ordinateurs double tous les 18 mois »[6].

Pour ces raisons nous allons tenter dans ce travail d'exposer, à travers le service Retrieve and Rank de Watson IBM, les composant essentiels qui permettent d'interroger une base de documents, chaque composant sera traité séparément pour essayer de montrer les possibilités et l'impact de chacun.

Pour finir, une analyse des résultats sera effectuée d'abord pour voir si une évolution est remarquée par rapport aux recherches classiques, ensuite pour tenter de voir comment le système adapte ses réponses par rapport à la façon dont nous l'entraînons, et enfin nous allons tenter de faire quelques conclusions.

Les tests porteront exclusivement sur un corpus d'environ 300 documents de lieux touristiques en Suisse issus d'internet[7]. Ce n'est pas un choix au hasard car avant de commencer mes études à la HEG, je travaillais comme créateur de voyages dans une agence de voyage. Selon mon expérience, nous pouvons gagner du temps en recourant à un système intelligent — dont le principe de fonctionnement consiste à prendre en compte les modèles des demandes traitées dans le passé — qui aide à traiter (ou à prétraiter) les demandes des clients, pour les adapter à leurs besoins spécifiques.

[6] https://fr.wikipedia.org/wiki/Loi_de_Moore

[7] Pour la plupart des commentaires de spécialistes ou d'amateurs de voyages sur le site http://www.monnuage.fr et http://www.cityzeum.com/

II. APPRENTISSAGE D'ORDONNANCEMENT

Parmi les techniques de *machine learning* utilisées pour les recherches textuelles, nous retrouvons les algorithmes d'apprentissage d'ordonnancement (learning to rank). Ces algorithmes — utilisés par le service Watson Retrieve and Rank — ont pour objectif d'ordonner les documents retrouvés en réponse à une requête ; ils utilisent un ensemble de données — appelé *Training data* — pour augmenter la pertinence des recherches.

En effet, dans ce type de recherches, chaque couple requête/document est transformé en un vecteur de caractéristiques[8] qui sont des propriétés mesurables représentant les scores de similarité entre la requête et le document. Nous pouvons également trouver parmi ces caractéristiques ceux qui sont propres à la requête ou au document comme, par exemple, le nombre de termes.

Le Training data, qui est constitué de vecteurs de caractéristiques, sera utilisé par les algorithmes pour créer des modèles d'ordonnancement (Ranking Model). Par la suite, lesdits modèles pourront permettre au système de prédire l'ordre optimal des documents pour une requête ultérieure.

Voici une architecture[9] possible d'un moteur de recherche utilisant l'apprentissage d'ordonnancement :

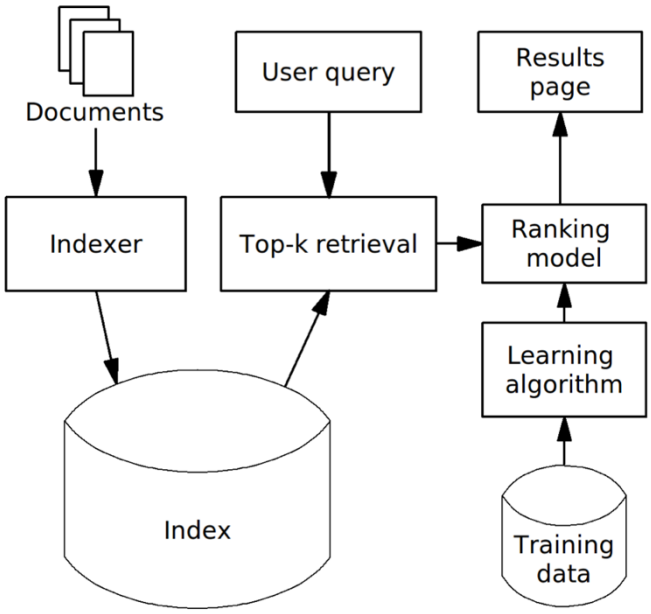


FIGURE 2: ARCHITECTURE POSSIBLE D'UN SYSTEME UTILISANT L'APPRENTISSAGE D'ORDONNANCEMENT

[8] http://thesesups.ups-tlse.fr/2170/1/2013TOU30240.pdf

[9] https://en.wikipedia.org/wiki/Learning_to_rank

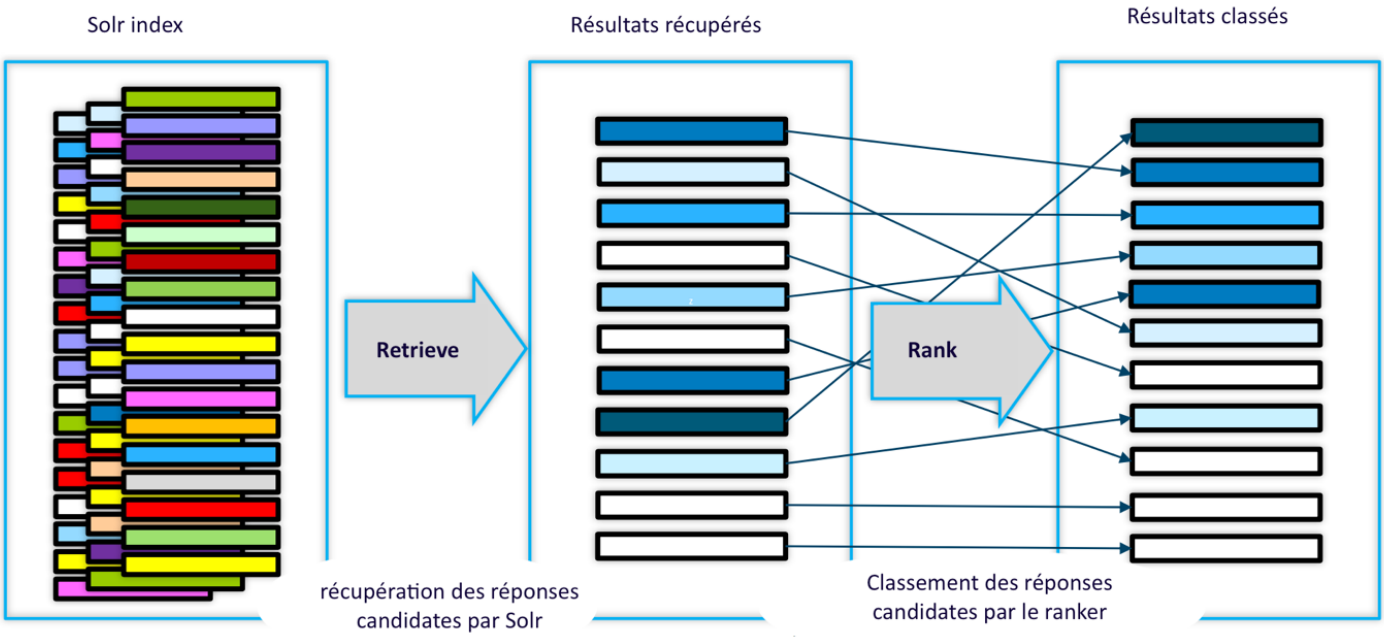


FIGURE 3 ARCHITECTURE GENERALE DE RETRIEVE AND RANK

Retrieve & rank (comme le montre la figure 3 [10])est un service Watson dont l'objectif est de trouver le document le plus pertinent parmi une collection de documents et ce, grâce aux technologies d'apprentissage profond (deep learning) mentionnées plus haut. Ainsi, le service R&R améliore la pertinence des documents retrouvés par rapport aux recherches classiques.

Les deux composants qui font la force de R&R sont Apache Solr et les technologies d'apprentissage profond.

S'agissant d'Apache Solr, nous commençons par préparer la collection de documents qui constituera une sorte de base de données dans laquelle le service va chercher pour répondre aux requêtes. Ensuite, nous devons créer un espace(cluster) qui va contenir notre collection en sachant que le cluster peut en contenir plusieurs.

Chaque collection représente un ensemble de sections contenant les données utilisées lors des recherches. Ces données vont être formatées de sorte que la

[10] Hiroaki Komine, slide 56, https://www.slideshare.net/komine/watson-api-20160716-rev02

signification de chaque partie soit connue de Solr, chaque partie pouvant être comparée aux colonnes des tableaux SQL.

Après avoir créé le cluster, la collection et les documents, le service devient capable de faire des recherches classiques.

La Ranker est le deuxième composant du service R&R, il peut être comparé à un enfant qui apprend au fur et à mesure des interactions avec le monde qui l'entoure. Pour apprendre au Ranker les réponses souhaitées, nous lui soumettons des exemples de questions qu'il pourrait recevoir et une note de pertinence sur chaque couple de question/document. Toutes ces données sont contenues dans un fichier dénommé le « Ground truth ».

Ground truth est utilisé par le système pour s'entrainer, et au fur et à mesure il devient de plus en plus pertinent dans ses recherches.

Voici le schéma[11] de la figure 4 qui résume les étapes d'implémentation du service R&R:

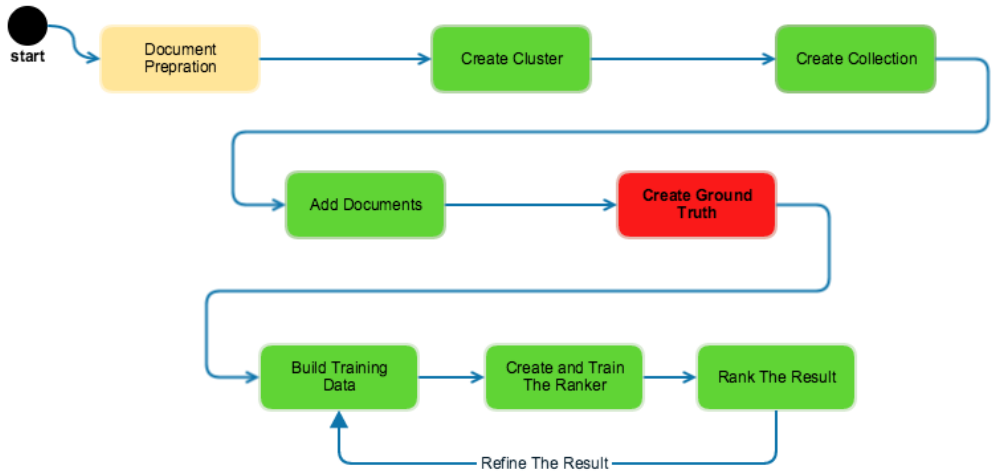


FIGURE 4: EXEMPLE DES ÉTAPES D'IMPLEMENTATION DU SERVICE R&R

[11] Fartash Haghani, http://fartashh.github.io/post/qa-system-watson/

A. DOMAINES D'UTILISATION

R&R peut s'avérer utile pour les services clients tel que les équipes de support, les techniciens et tous les métiers qui ont besoin de trouver rapidement l'information la plus pertinente parmi une grande masse de documents.

Par exemple, R&R peut s'avérer utile pour un technicien souhaitant trouver une solution dans un manuel d'utilisation très volumineux. Tout comme ce service peut s'avérer utile pour les services professionnels en chasse des personnes talentueuses possédant les compétences appropriées.

Dans les forums tels que « stackoverflow », il est des questions récurrentes de sorte que la plupart des visiteurs de cette plateforme recherchent d'abord parmi les questions déjà posées, et donc les réponses apportées. Malheureusement cette recherche peut s'avérer difficile car la recherche, classique, implémentée par les moteurs de recherche (google, yahoo, etc…) recourt aux mots clés. C'est là qu'intervient la plue-value de R&R qui peut améliorer cette recherche en utilisant son système de *ranking*, car ladite recherche peut comprendre le sens du mot pour fournir les réponses existantes les plus pertinentes.

B. NOTRE CAS D'ÉTUDE

La performance d'une application utilisant le service R&R peut être mesurée à sa capacité de retourner les documents les plus pertinents pour l'utilisateur.

Afin d'expérimenter le service R&R, nous avons décidé de créer un système qui contient un ensemble de lieux touristiques (des restaurants, musées, monuments etc…), et qui retournera une liste des lieux les plus appropriés par rapport à une requête.

L'objectif étant de vérifier la capacité du système à retourner les documents les plus pertinents, nous avons observé l'évolution de la pertinence des réponses suivant le nombre et le type de questions.

C. AVANT DE COMMENCER

La première étape avant de commencer est de créer un compte bluemix pour pouvoir utiliser les services offerts par Watson. Ensuite il faut créer une instance du service Retrieve and Rank pour avoir les *credentials* (*username* et *password*), en effet pour chaque instance d'un service Watson nous avons besoin d'un accès.

En résumé nous avons eu besoin pour ce travail de :

- Créer un compte Bluemix[12]
- Créer une instance du service Retrieve and Rank
- Créer une instance du service Documents conversion

Pour chacune de ces sections, nous avons un *username* et un *password*.

Dans l'annexe 1 vous trouverez les étapes pour obtenir les *credentials* ainsi qu'une capture d'écran de ces derniers.

D. RETRIEVE

1. SOLR

Le composant « Retrieve » du service Retrieve & Rank est implémenté par Apache Solr qui est l'un des outils de recherche Open Source les plus répandus (avec Sphinx et ElasticSearch).

Pour effectuer ses recherches, Solr conserve les données sous un certain format. Ainsi, chaque document (appelée également section) est défini comme un ensemble de champs (fields) auxquels sont associées des valeurs. Ces champs permettent à Solr d'indexer les données et les métadonnées pour pouvoir effectuer les recherches idoines.

Comme mentionné précédemment, la partie *Retrieve* du service R&R est basée sur Apach Solr, de sorte que quelques connaissances du fonctionnement de Solr sont nécessaires car bien souvent R&R transmet les messages de Solr sans modification à l'utilisateur. Par exemple, lorsque Solr retourne une erreur au service R&R, celui-ci transmet cette erreur sans modification à l'utilisateur.

Normalement, la préparation des documents est la première étape. Toutefois, pour comprendre pourquoi les documents non structurés sont formatés d'une certaine façon, nous avons commencé par décrire la configuration de Solr.

2. CONFIGURATION

Solr donne la possibilité d'être adapté et configuré selon nos besoins. Afin de faciliter la tâche, IBM Bluemix a fourni un dossier de configuration Solr sur lequel nous nous somme appuyés pour paramétrer les recherches Solr selon les spécificités de notre cas qui traite des requêtes en Français.

[12] https://console.bluemix.net/

Nous allons d'abord expliquer les parties qui concernent les recherches du Service R&R, pour ensuite adapter aux spécificités de notre cas.

La structure du dossier de configuration est la suivante :

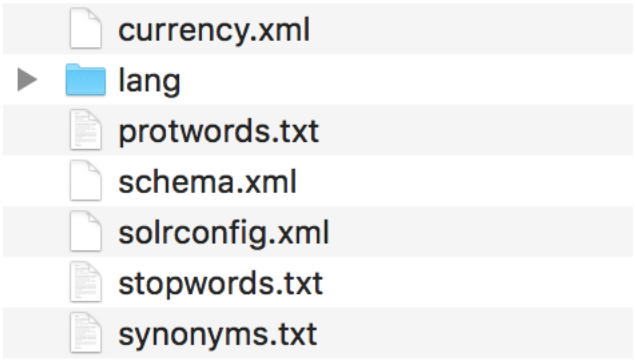


FIGURE 5: STRUCTURE DU DOCUMENT DE LA CONFIGURATION SOLR

Globalement les deux fichiers XML paramétrables qui nous intéressent sont :

- Solrconfig.xml
- Schema.xml

A) SOLRCONFIG.XML

Solrconfig.xml [13] concerne la configuration au niveau système de Solr. Il est par exemple possible de modifier la quantité de mémoire RAM attribuée à Solr. Compte tenu de la faible incidence sur notre cas d'étude quant aux modifications apportées à ce fichier, nous n'allons pas nous attarder sur cette partie.

B) SCHEMA.XML[14]

Le schéma de l'index (Schema.xml[15]) donne la liste de tous les champs d'un document Solr (partie 1 de la figure), chaque élément de la liste peut comprendre de nombreuses options parmi lesquelles : le type du champ, la possibilité de calculer la valeur d'un champ à partir d'un autre ainsi que des traitements divers à appliquer à la valeur du champ, etc.

[13] http://igm.univ-mlv.fr/~dr/XPOSE2008/Apache%20Solr/solr_configuration.html

[14] Nicolas Travers, Raphaël Fournier S'niehotta et Philippe Rigaux , 2014-2017, http://b3d.bdpedia.fr/solr.html

[15] http://b3d.bdpedia.fr/solr.html

Dans ce fichier, nous retrouvons également les types de champs, ainsi que leurs comportements (partie 3 de la figure). Pour faire la liaison entre le type et son comportement chaque type mentionné dans les champs doit apparaître dans un des éléments fieldType du fichier schema.xml(la flèche blanche).

Voici le squelette[16] des parties les plus importantes du fichier schema.xml *:*

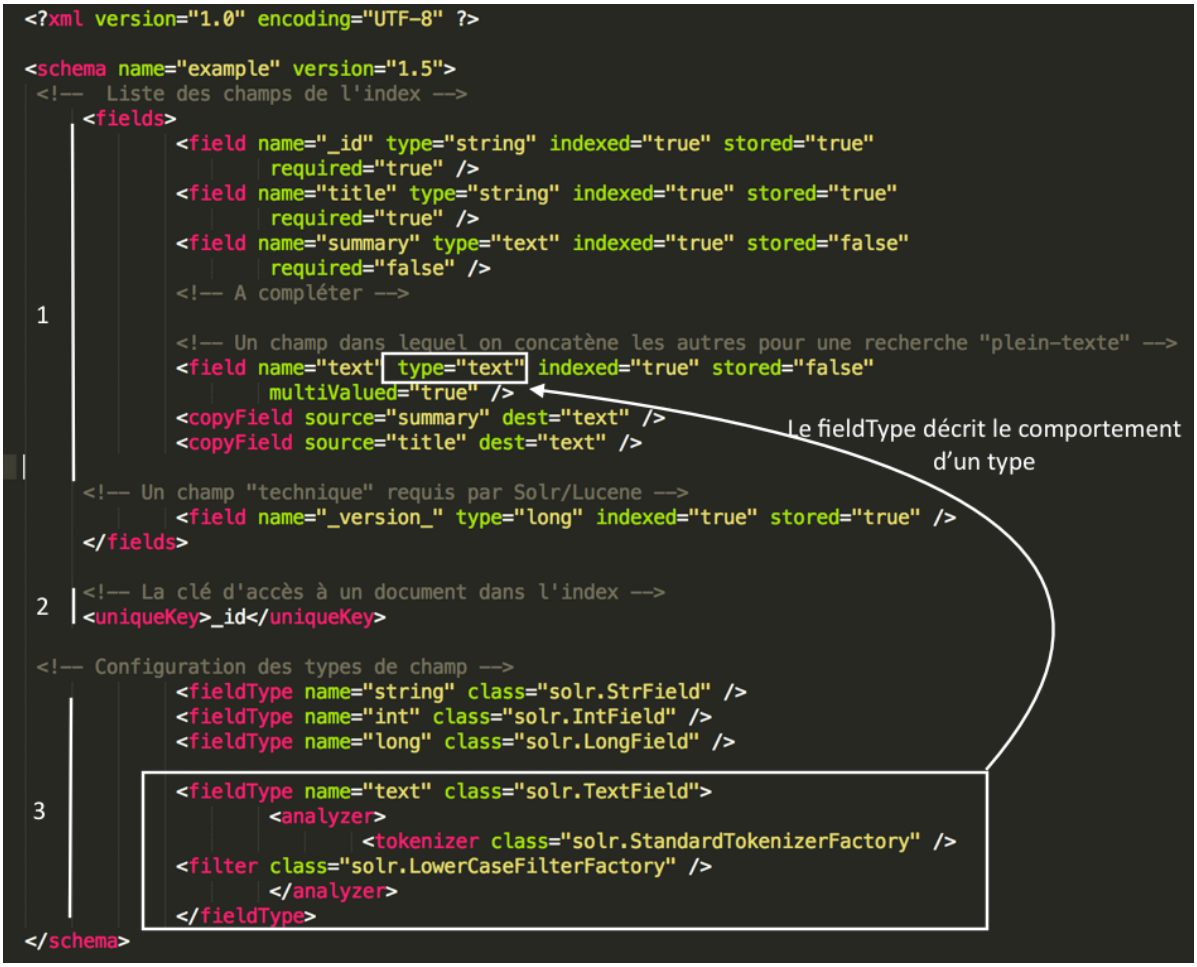


FIGURE 6: SQUELETTE DU SCHEMA.XML

En bref, le fichier Schema.xml comprend trois parties :

1) La liste des champs, dans l'élément fields

2) Un champ qui unique qui identifie le document;

3) La liste des types de champs (ou *fieldTypes).*

[16] http://b3d.bdpedia.fr/solr.html

C) LES TYPES SOLR

Solr fournit tout un ensemble de types pré-définis. En effet un type Solr peut être comparé à une classe Java avec des types primitifs (int, double etc…), des types spécifiques à Solr comme par exemple le type text, ainsi que des classes dont nous pouvons modifier le comportement. Les types spécifiques à Solr sont toutefois très généraux et donc pas nécessairement adaptables aux spécificités de chaque cas.

Solr donne la possibilité de créer des types spécifiques auxquels nous pouvons associer des options qui peuvent représenter, pour Solr, la façon dont le filtrage et la tokenization ainsi que d'autres techniques vont être appliqués à un certain champ lors de la recherche ou de l'indexation. C'est donc au développeur de définir ce dont son application a besoin comme options pour effectuer les recherches.

Pour comprendre l'objectif de ces techniques (filtrage, tokenization, etc..), voici quelques exemples[17] :

- *Si nous cherchons les documents contenant le mot "loup", on s'attend à trouver ceux contenant "loups", "Loup", "louve"; pour cela nous devons supprimer les majuscules et les pluriels; le cas de "loup / louve" illustre une transformation qui dépend de la langue;*
- *Un tokenizer prend en entrée un texte (une chaîne de caractères) et produit une séquence de tokens. Il effectue donc un traitement purement lexical, consistant typiquement à éliminer les espaces blancs, la ponctuation, les liaisons, etc., et à identifier les "mots". Des transformations peuvent également intervenir (suppression des accents par exemple, ou normalisation des acronymes, —U.S.A. devient USA).*

Pour résumer, voici les éléments[18] essentiels lors de l'analyse :

- Tokenization: découpage du texte en "termes"(selon la ponctuation, les espaces, etc),
- Normalisation : identification de toutes les variantes d'écritures d'un même terme et choix d'une règle de normalisation (que faire des majuscules ? Acronymes ? Apostrophes ? Accents ?).
- *Stemming* ("racinisation") : rendre la racine des mots pour éviter le biais des variations autour d'un même sens (auditer, auditeur, audition, etc.)
- Stop words ("mots vides") : les Stop words sont des mots qui n'apportent pas grand chose aux recherches. Ces termes ou mots vides sont trop commun ou

[17] Ces exemples proviennent du site internet http://b3d.bdpedia.fr/solr.html

[18] http://b3d.bdpedia.fr/solr.html

trop nombreux. En principe il est inutile de les indexer, pour cela il est important de les exclure complétement de la recherche.

Examinons[19] maintenant la définition d'un champ dans le fichier de configuration Schema.xml, avec l'exemple de l'identifiant.

```
<field name="id" type="string" indexed="true" stored="true" required="true" multiValued="false" />
```

Les attributs de l'élément XML caractérisent le champ. Le nom et le type sont les informations de base. Ensuite, nous pouvons trouver toutes sortes d'attributs. La plupart, ayant une valeur par défaut, sont optionnels. Voici quelques exemples d'options :

- *indexed indique simplement que le champ peut être utilisé dans une recherche;*
- *stored indique que la valeur du champ est stockée dans l'index, et qu'il est donc possible de récupérer cette valeur comme résultat d'une recherche, sans avoir besoin de retourner à la base principale; en d'autres termes, stored permet de traiter l'index aussi bien comme une base de données;*
- *required indique que le champ est obligatoire;*
- *enfin, multivalued vaut true pour les champs ayant plusieurs valeurs, soit, concrètement, un tableau en JSON; c'est le cas par exemple pour le nom des acteurs.*

Enfin, dans le fichier schema.xml de la configuration fournie par Watson, nous retrouvons un type nommé watson_text_en, configuré pour être inclus dans le *R&R query parser* pour les requêtes de R&R en anglais. Ainsi donc, le *query parser* est utilisé pour analyser les requêtes et pour en faire par la suite une requête compréhensible par Apach Solr.

D) ADAPTER LE TYPE AUX REQUÊTES EN FRANÇAIS

Qu'il s'agisse du Français ou de l'Anglais, la tokenisation (diviser le texte en *tokens* ("mots")) est très fortement dépendante de la langue, La racinisation est également très dépendante de la langue et peut nécessiter une analyse linguistique complexe. En anglais, *geese* est le pluriel de *goose* (oie), *mice* de *mouse* (souris) ; les formes masculin / féminin en français n'ont parfois rien à voir ("taureau / vache") voici encore quelques exemples[20] :

[19] Ces exemples proviennent du site internet http://b3d.bdpedia.fr/solr.html

[20] Ces exemples proviennent du site internet http://b3d.bdpedia.fr/solr.html

- *Dans certaines langues (Chinois, Japonais), les mots ne sont pas séparés par des espaces.*
- *Certaines langues s'écrivent de droite à gauche, de haut en bas.*
- *Que faire (et de manière cohérente) des acronymes, élisions, nombres, unités, URL, email, etc.*
- *Que faire des mots composés : les séparer en tokens ou les regrouper en un seul ? Par exemple :*
  - *Anglais : hostname, host-name et host name, …*
  - *Français: Le Mans, aujourd'hui, pomme de terre, …*
  - *Allemand: Levensversicherungsgesellschaftsangestellter (employé d'une société d'assurance vie).*

Voici la liste des langues prises en charge par le service R&R:

Language	Tokens	Lemmas
English (`en`)	Yes	Yes
Spanish (`es`)	Yes	Yes
Brazilian Portuguese (`ptbr`)	Yes	No
Japanese (`ja`)	Yes	No
Arabic (`ar`)	Yes	No
French (`fr`)	Yes	No
Italian (`it`)	Yes	No
German (`de`)	Yes	No

FIGURE 7: LANGUES PRISES EN CHARGE PAR LE SERVICE R&R

La qualité des résultats peut changer selon la langue. En effet pour la langue française la tokenisation est prise en charge par le *query parser* par contre la lemmatisation ou racinisation ne l'est pas encore.

Afin d'adapter les recherches aux spécificités de la langue française nous avons créé un nouveau type que nous avons nommé watson_text_fr :

```
<fieldType name="watson_text_fr" indexed="true" stored="true" omitNorms="false" omitTermFreqAndPositions="false"
storeOffsetsWithPositions="true" class="solr.TextField">
  <analyzer type="index">
    <tokenizer class="com.ibm.watson.search.analysis.sire.SireTokenizerFactory" tokens="true" lemmas="false" language="fr"/>
      <filter class="solr.LowerCaseFilterFactory"/>
  </analyzer>
  <analyzer type="query">
      <tokenizer class="com.ibm.watson.search.analysis.sire.SireTokenizerFactory" tokens="true" lemmas="false" language="fr"/>
      <filter class="solr.LowerCaseFilterFactory"/>
      <filter class="solr.StopFilterFactory"
      ignoreCase="true"
      words="lang/stopwords_fr.txt"
      />
  </analyzer>
</fieldType>
```

Dans le fieldType, où nous trouvons la liste des types de champ, nous remarquerons que[21] :

- class=solr.TextField
  En effet ce sont des types configurables qui surpassent de très loin les fonctionnalités offertes par les opérateurs LIKE et REGEXP du langage SQL, en effet c'est un type complexe qui est configuré à partir du type solr.TextField auquel nous pouvons ajouter un ou deux analyseurs (<analyzer></analyzer>), par exemple pour décrire le comportement lors de l'indexation ainsi que lors de la recherche, ces analyseurs sont eux-mêmes composés d'une suite de <tokenizer/> et de <filter/>.
- language =fr indique à Solr que la recherche sera effectuées en français,
- La tokenisation est activée,
- La lemmas(racinisation) est désactivée car non disponible pour la langue française, à noter que si la valeur de lemmas est *true*, cela provoquera une exception,
- omitNorms="false" car nous n'avons pas besoin de trier sur ce champ,
- omitTermFreqAndPositions="false" car nous n'avons pas besoin ni de la fréquence des termes ni de leur position ni de leur pondération (*payload*)
- storeOffsetsWithPositions="true",

  A noter que si nous ne mentionnons pas de valeurs pour les tokens et les lemmas, le service prend en compte la tokenisation — existante dans toutes les langues — mais pas la lemmatization,
- Les stopwords [22] . Nous avons ajouté dans la configuration le fichiers stopwords_fr.txt pour les requêtes en Français.

Etant donné que nous avons défini un type personnalisé, nous devons l'inclure dans le fichier solrconfig.xml pour le gestionnaire de requêtes fcselect car c'est ce dernier qui fournit la fonctionnalité de classement pour le service R&R.

```
<queryParser name="fcQueryParser" class="com.ibm.watson.hector.plugins.qparser.WatsonFCQParser">
    <str name="textFieldTypes">watson_text_fr</str>
</queryParser>
```

[21] http://g-rossolini.developpez.com/tutoriels/solr/?page=schema

[22] Le fichier stopwords_fr.txt provient du site : http://referencement-gratuit.and-co.ch/liste-stopwords-francais/

En effet, lorsque watson_text_fr est mentionné dans cette partie du solrconfig.xml, le service R&R sait qu'il doit donner un classement aux documents retournés lors d'une requête fcselect.

A noter que lorsque nous ajoutons le paramètre fcselect à une requête de recherche, le service R&R sait qu'il doit retourner les résultats en utilisant le Ranker que nous allons décrire plus bas.

E) DEFINITION DES CHAMPS

Après avoir adapté le type de champs aux types de recherches que nous allons effectuer, nous allons définir les champs dont aurons besoin pour notre cas. Les voici donc :

- id — l'identifiant unique du document(ou section) ;
- title — titre du document ;
- content_text — le text du document.

Dans la liste de champs comme dans le fichier Schema.xml :

```
-->
<field name="id" type="string" indexed="true" stored="true" required="true" multiValued="false" />
<field name="title" type="watson_text_fr" indexed="true" stored="true" required="false" multiValued="true" />
<field name="content_text" type="watson_text_fr" indexed="true" stored="true" required="false" multiValued="true" />
```

Maintenant que nous connaissons la structure des documents dont Solr a besoin pour faire ses recherches, nous avons besoin de formater les données non structurées dans le format Solr que nous venons de créer dans le fichier de configuration Schema.xml. Pour cela nous allons utiliser le service Document Conversion fournit par IBM. Mais nous allons d'abord préparer les documents dans lesquels le service R&R va chercher ceux qui répondent aux requêtes des utilisateurs.

Ce filtrage permet de garder seulement les informations dont nous avons besoin. Par exemple nous pouvons exclure un type de balises, ce qui permet de n'avoir que les informations pertinentes dans les sections que nous allons transmettre au service R&R.

3. DOCUMENTS CONVERSION

Le service Document Conversion est un service Watson qui peut convertir des données non structurées en données normalisées en format HTML, format Plain text ou en format JSON appelé « Answers Units », ces derniers peuvent être intégrés avec d'autres services Watson -comme R&R-, la figure 8 montre un exemple de cas d'utilisation :

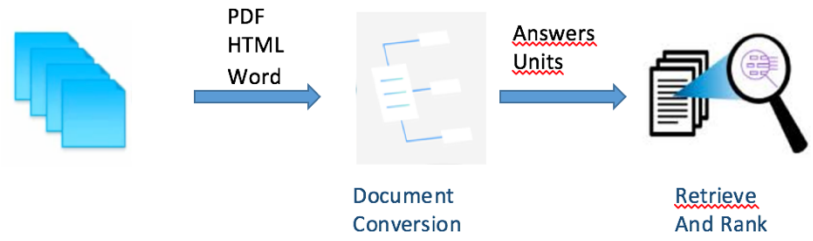


FIGURE 8: INTEGRATION DU SERVICE DOCUMENT CONVERSION AVEC LE SERVICE R&R

A) ANSWERS UNITS

Les Answers Units sont des données formatées en JSON de la façon suivante[23] :

```
"answer_units": [
  {
    "id": "276a84a7-348c-413c-bbd6-eed180257694",
    "type": "h3",
    "parent_id": "",
    "title": "What is Watson?",
    "direction": "ltr",
    "content": [
      {
        "media_type": "text/plain",
        "text": "Watson is an artificially intelligent computer system capable of answering
      }
    ]
  },
```

Dans cet exemple, nous retrouvons les données et les méta-données suivantes :

- *id* : l'*id* de la section
- *type* : h3 qui représente le type du titre, h3 peut représenter la balise en html ou niveau du titre en word ou pdf comme nous allons voir plus bas,
- *title* : le titre de la section
- *content* : qui contient à son tour le texte et le media_type du texte qui est

[23] L'exemple provient de la documentation du site IBM bluemix

dans ce cas un text/plain.

En effet la conversion passe par plusieurs étapes, comme le montre la figure 9[24]. D'abord le service commence par convertir le ficher original (word ou pdf) en html (si le fichier est déjà un HTML, cette étape est ignorée). Ensuite, il le convertit en un format HTML normalisé. Cette phase de normalisation nettoie les balises HTML qui ne contiennent pas de données utiles, comme les bas de pages. Par la suite si le format demandé est JSON, *Answers Units* ou *plain text, le service* convertit encore au format voulu.

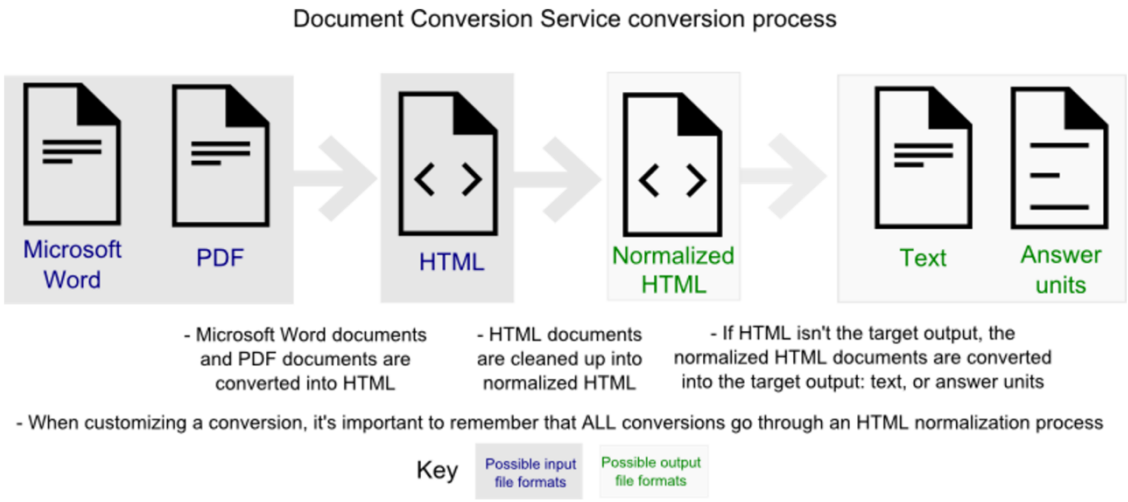


FIGURE 9: PROCESSUS DE CONVERSION DU SERVICE DOCUMENT CONVERSION

Nous pouvons *customiser* la conversion avec un text JSON que nous transmettons au service Document Conversion, ce JSON contient :

(1) LE FORMAT VISÉ

normalized_html, plain text ou answer_units.

Par exemple si nous désirons avoir des answer units (JSON) :

```
"conversion_target": "answer_units",
```

[24] https://console.bluemix.net/docs/services/document-conversion/index.html

(2) PARAMETRAGE

La façon dont il va identifier les sections dans le document. En effet le service permet de nettoyer le contenu non désiré, en ne gardant que les titres qui sont d'un niveau hiérarchique[25] de balise pour html et de titre pour PDF et Word.

Pour formater des données en Answers Units, le service donne une grande marge de manœuvre pour paramétrer le fichier en entrée et le fichier en sortie, par exemple pour un document Word, nous avons la possibilité, comme le montre l'exemple suivant, de définir le niveau (*level*), le min_size, le max_size ainsi que le type d'écriture qui seront garder lors de la conversion :

```
{
  "word": {
    "heading": {
      "fonts": [
        {"level": 1, "min_size": 24},
        {"level": 2, "min_size": 18, "max_size": 23, "bold": true},
        {"level": 3, "min_size": 14, "max_size": 17, "italic": false},
        {"level": 4, "min_size": 12, "max_size": 13, "name": "Times New Roman"}
      ],
      "styles": [
        {
          "level": 1,
          "names": ["pullout heading", "pulloutheading", "heading"]
        },
        {
          "level": 2,
          "names": ["subtitle"]
        }
      ]
    }
  }
}
```

La description des différents champs est dans les annexes (annexe 2).
Ensuite le service permet de définir des options supplémentaires pour raffiner le nettoyage des documents en entrée(input), ce qui constitue la phase de normalisation décrite plus haut — passer d'un fichier HTML à un fichier HTML normalisé —, voici un exemple de normalisation qui permet de supprimer le *tag*, comme le *tag strong* tout en gardant le contenu :

[25] Exemple de niveau hiérarchique en HTML : h1, h2, h3 etc…

```
"normalized_html": {
    "exclude_tags_keep_content": ["font", "em", "span", "strong", "code"]
  },
```

La liste complète des différents champs se trouve dans les annexes (annexe 3)

Dans le cas où notre target_content ou fichier de sortie contient des Answer Units, nous devons indiquer au service les balises de titre (h1, h2, h3, etc) qui vont devenir le titre de chaque Answer Unit en sachant que tout ce qui est plus bas (hiérarchiquement dans le document) va devenir le content_text de l'Answer Unit.

```
{
"answer_units": {
  "selector_tags": ["h1", "h2", "h3", "h4", "h5", "h6"]
}
```

Si nous ne définissons pas de selector_tags, le service prend pas défaut h1 et h2.

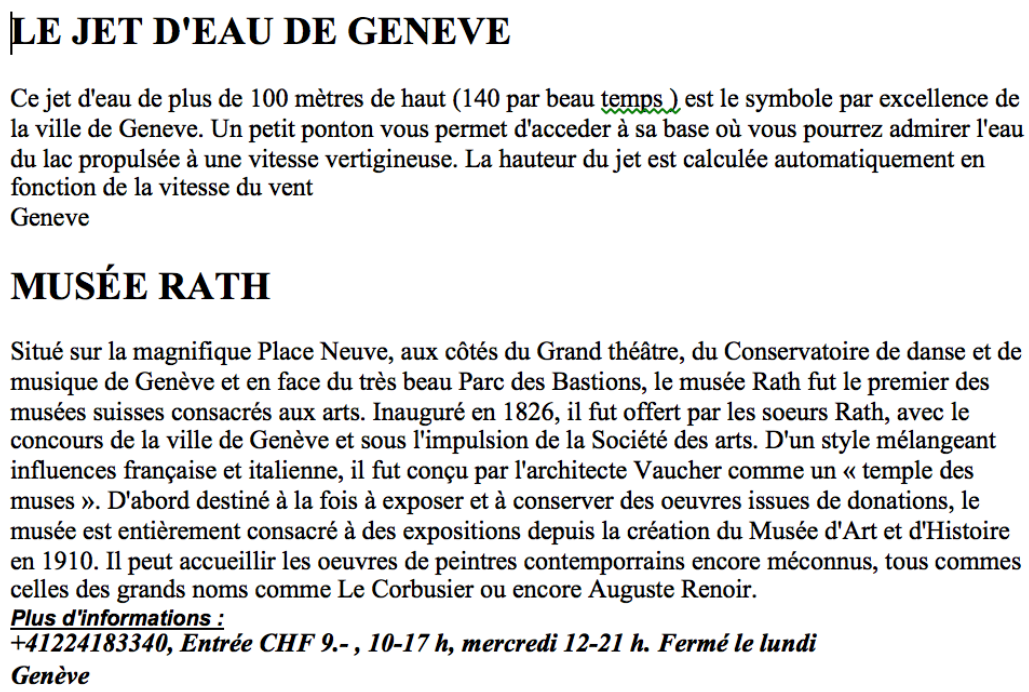
En mettant tout les éléments ensemble[26] :

```
{
  "conversion_target": "answer_units",
  "word": {
    "heading": {
      "fonts": [
        {"level": 1, "min_size": 24},
        {"level": 2, "min_size": 18, "max_size": 23, "bold": true},
        {"level": 3, "min_size": 14, "max_size": 17, "italic": false},
        {"level": 4, "min_size": 12, "max_size": 13, "name": "Times New Roman"}
      ],
      "styles": [
        {
          "level": 1,
          "names": ["pullout heading", "pulloutheading", "heading"]
        },
        {
          "level": 2,
          "names": ["subtitle"]
        }
      ]
    }
  },
  "normalized_html": {
    "exclude_tags_keep_content": ["font", "em", "span", "strong", "code"]
  },
  "answer_units": {
    "selector_tags": ["h1", "h2", "h3", "h4", "h5", "h6"]
  }
}
```

4. PREPARATION DES DOCUMENTS

Dans notre exemple nous avons créé un ensemble de documents à partir des descriptions fournies par des visiteurs sur certains lieux touristiques (restaurants, musées, monuments, etc). En premier lieu, une phase de collecte est effectuée à la main pour mettre les données dans un fichier Word, chaque lieu sera considéré comme une section avec le nom du lieu et sa description. Cette structure a pour objectif de faciliter au service Document Conversion le formatage du document non structuré dans le format « Answer Unit » décrit plus haut. Prenons par exemple :

[26] https://console.bluemix.net/docs/services/document-conversion/index.html



LE JET D'EAU DE GENEVE

Ce jet d'eau de plus de 100 mètres de haut (140 par beau temps) est le symbole par excellence de la ville de Geneve. Un petit ponton vous permet d'acceder à sa base où vous pourrez admirer l'eau du lac propulsée à une vitesse vertigineuse. La hauteur du jet est calculée automatiquement en fonction de la vitesse du vent
Geneve

MUSÉE RATH

Situé sur la magnifique Place Neuve, aux côtés du Grand théâtre, du Conservatoire de danse et de musique de Genève et en face du très beau Parc des Bastions, le musée Rath fut le premier des musées suisses consacrés aux arts. Inauguré en 1826, il fut offert par les soeurs Rath, avec le concours de la ville de Genève et sous l'impulsion de la Société des arts. D'un style mélangeant influences française et italienne, il fut conçu par l'architecte Vaucher comme un « temple des muses ». D'abord destiné à la fois à exposer et à conserver des oeuvres issues de donations, le musée est entièrement consacré à des expositions depuis la création du Musée d'Art et d'Histoire en 1910. Il peut accueillir les oeuvres de peintres contemporrains encore méconnus, tous commes celles des grands noms comme Le Corbusier ou encore Auguste Renoir.
Plus d'informations :
+41224183340, Entrée CHF 9.- , 10-17 h, mercredi 12-21 h. Fermé le lundi
Genève

MUR DES RÉFORMATEURS

Situé au coeur du Parc des Bastions, en plein centre de Genève, le Mur des Réformateurs est un lieu incontournable de la ville, capitale du protestantisme par excellence. Créé en 1909, année du 400e anniversaire de la naissance de Jean Calvin (l'un des initiateurs de la Réforme protestante) et année du 350e anniversaire de l'Académie de Genève (ancêtre de l'université), le mur est adossé à l'une des murailles qui entourent la colline de la vieille-ville de Genève. Les quatre grandes figures du mouvement protestant sont représentées au centre du mur et ont une hauteur

FIGURE 10 STRUCTURE DU DOCUMENT WORD UTILISÉ

Vu que dans notre cas nous avions déjà nettoyé les documents à la main, en mettant les données dans un fichier Word et en mettant le titre de chaque lieu en « Titre 2 », voici la configuration de paramétrage JSON choisie :

```
{"answer_units":
  {"heading":
    {"fonts":[{"level":2,"min_size":1}]
    }
  }
}
```

Cette configuration demande en sortie des Answer Units, en ne gardant dès la première phase de transformation du texte en html (Phase 1 dans le processus de Document Conversion) que les « Titre 2 », avec le min_size égale à 1. Nous n'avons

pas défini de selector_tags car le nettoyage a été fait à la phase 1.

La figure suivante est un exemple de l'*output* de document conversion en format JSON où chaque partie représente un Answer Unit :


```
{
  "content": [
    {
      "media_type": "text/plain",
      "text": "julie hattu: L'endroit à ne pas louper ! Boire un verre à 105 m de haut,
      à Bâle c'est possible ! Le café rouge est situé au 31ème étage de la tour de la foire
      qui est la tour la plus haute de Suisse ! Quand on sort de l'ascenseur, la vue sur toute
      la ville est époustouflante, de quoi donner le vertige ! L'ambiance y est plutôt lounge,
      des grands canapés permettent de se prélacer devant la vue tout en buvant un verre. C'est
      bien agréable après une journée de marche touristique. Un point fort de cet endroit qui fera
      plaisir à certains, il y a un fumoir au centre du bar qui permet de profiter aussi de la vue
      et évite de descendre et monter les 31 étages ! Enfin, le top du top, les toilettes ! Hommes
      comme femmes l'expérience vaut le coup d'œil ! mcfly: L'un des bars les plus hauts d'Europe,
      situé au 31ème (et dernier) étage du plus haut building de Bâle. ☎ 41 613 613 031 - Level 31 Messeturm, Bâle"
    }
  ],
  "id": "b733604f-ac1e-4fa5-b281-072f43a535a4",
  "title": "Bar Rouge",
  "type": "h2",
  "direction": "ltr",
  "parent_id": ""
},
{
  "content": [
    {
      "media_type": "text/plain",
      "text": "Il y a des musées qui te procurent des sensations beaucoup
      plus inattendues de ceux à quoi tu t'attends. C'est ce qui m'est arrivé
      dans cet étrange musée du papier. Situé dans un ancien moulin, ce lieu imprégnés
      d'histoire a été la source d'importante fabrication de livre ces derniers siècles.
      À travers ses différentes salles, le musée nous offre un aperçu de ce que pouvez être
      l'histoire du papier et de l'imprimerie. Ce qui est génial c'est qu'il te laisse t'immerger
      dans chacune de ces étapes. C'est à dire qu'au rez-de-chaussée, tu peux te mettre dans la peau
      d'un fabricant de papier et apprendre à utiliser le matériel traditionnel. Ceci te permettra de
      créer ta propre feuille. Une fois la feuille fabriquait, tu passes à l'étape suivante qui se trouve
      à l'étage supérieur. Cette étape consiste à écrire ou inscrire différent élément sur le papier.
      Franchement c'est un endroit très chouette qui sera encore mieux lorsque les travaux à l'extérieur s'arrêteront.
      De plus, il se situe dans un magnifique quartier, tous proche de la rivière"
    }
  ],
  "id": "9e7af9bf-cb07-4e4c-ba2e-c80c457bf3a2",
  "title": "Musée du Papier",
  "type": "h2",
  "direction": "ltr",
  "parent_id": ""
}
```


FIGURE 11: STRUCTURE DES ANSWERS UNITS RETOURNEE

Notez que nous avons séparé les lieux car le service R&R (plus spécifiquement Solr) a besoin d'avoir les informations sous une certaine structure pour pouvoir indexer les documents dans la collection et pour pouvoir effectuer des recherches.

Dans l'annexe 4 (partie 1), vous trouverez, avec la description de chaque partie, l'implémentation en code JAVA des étapes mentionnées plus haut.

5. CREATION DU SOLR CLUSTER

Après avoir adapter la configuration Solr et préparer les données non structurées et les avoir formater en Answers Units, maintenant nous sommes près pour la phase d'indexation des Answers Units dans Solr, mais avant cela nous devons créer un espace qui contient les documents, pour cela nous avons créé un espace Solr cluster.

Un Solr cluster est un espace qui contient les collections de recherche qu'on devrait créer plus tard, en effet, c'est dans les collections que nous poserons les documents.

Voir le code création du cluster avec la description de chaque partie du code dans l'annexe 5.

6. CREATION DE LA COLLECTION

Une collection est un ensemble de section dans lesquels on peut rechercher lors d'une requête. L'équivalent en SQL serait la table[27].

Chaque collection est associée à une configuration qui est constituée d'un ensemble de fichiers de configuration Solr comme mentionnés plus haut (shema.xml, solrconfig.xml, …). Une configuration peut être utilisée par plusieurs collections[28].

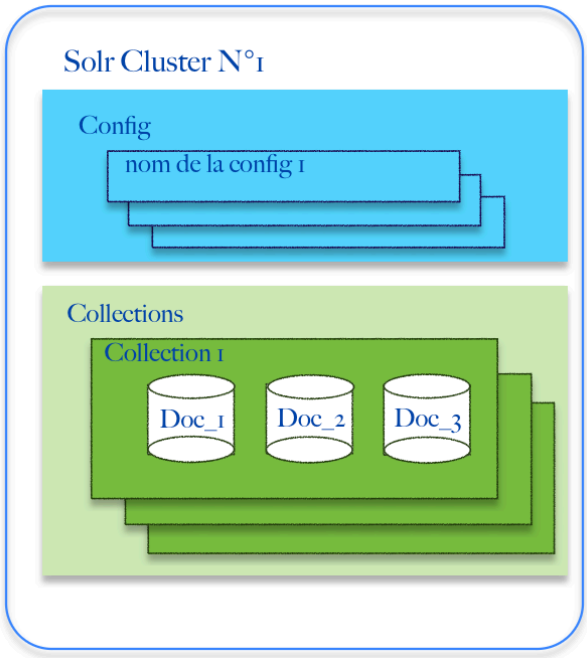


FIGURE 12: STRUCTURE DE LA PARTIE RETRIEVE

Dans cette phase nous allons créer la collection, et l'associer à notre configuration pour pouvoir charger les documents dedans par la suite.

Dans l'annexe 6 vous trouverez l'implémentation de cette étape, avec la description de chaque partie.

[27] https://www.eolya.fr/2013/05/30/presentation-de-solrcloud-solr-4-0/

[28] http://qiita.com/VegaSato/items/6d2d03d6a8b42adcf87e

7. INDEXER LES DOCUMENTS

Dans cette phase nous indexons les documents, soigneusement structurées, selon une structure adaptée aux champs de la configuration de la collection créée précédemment. En utilisant la liste des Answers Units retournées par document conversion nous allons itérer sur chacun de ces Answers Units pour mettre chaque élément dans le champ(*field*) qui convient.

NB : nous pouvons également dans cette phase contrôler le contenu de chaque Answer_Unit pour filtrer les informations que nous n'avons pas besoin, par exemple lorsque le titre est un chiffre.

L'implémentation de cette partie se trouve dans l'annexe 4 (partie 2).

E. RANK

1. CREER ET ENTRAINER LE RANKER

Après avoir posé la base pour les recherches classiques implémentées par Solr, nous passons à la deuxième partie qui consiste à utiliser les techniques d'apprentissage d'ordonnancement pour améliorer la pertinence des réponses. Le composant responsable de cette partie s'appelle le *Ranker*[29].

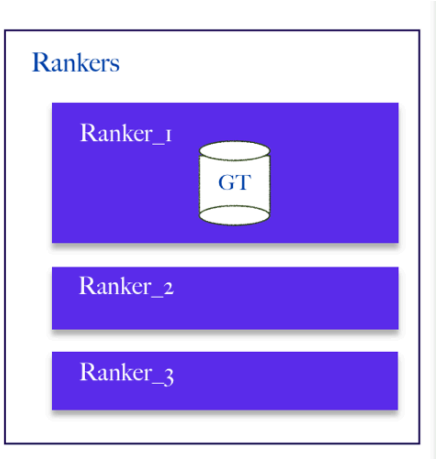


FIGURE 13: STRUCTURE DE LA PARTIE RANK

La force du Ranker provient de sa capacité d'apprendre à partir des exemples de questions et réponses que nous lui fournissons. Après cette phase d'apprentissage, le Ranker est sensé utiliser la même logique pour répondre aux requêtes qu'il n'a jamais rencontrées.

A) GROUND TRUTH

Afin d'entrainer Retrieve & Rank nous commençons par créer la « Ground truth ».

L'objectif de ce dernier est de permettre au service R&R d'extraire les caractéristiques(features), les plus importantes dans un couple question/réponse_candidate et les transformer en un vecteur de caractéristiques.

Dès lors que les données ont été transformées en vecteurs de caractéristiques, Watson utilise des formules statistiques pour créer des modèles lui permettant de classifier la pertinence d'un document quelconque.

Pour entrainer le service R&R, la procédure est la suivante :

[29] http://qiita.com/VegaSato/items/6d2d03d6a8b42adcf87e

1. Soumettre une requête à l'API du service R&R contenant la question avec laquelle nous avons l'intention d'entrainer le service, suivie des réponses possibles et une note sur la pertinence de chacune des réponses (relevance_label).
   Si le document répond complètement (c'est une réponse) 4*, si la réponse est incomplète c'est 3*, si c'est un document qui traite le sujet mais n'apporte pas de réponses c'est 2*, sinon 1*.

   Ces questions sont des questions typiques que le système peut recevoir d'un utilisateur.

   En effet voici les paramètres que l'API a besoin pour la création des vecteurs de caractéristiques (phase 2) :

```
q=in practice, how close to reality are the assumptions that the flow in a hypersonic shock tube using
&gt=656,2,1313,2,1317,2,1316,1,1318,2,1319,2,1157,2,1274,2,1286,0
&generateHeader=false
&rows=1
&returnRSInput=true
&wt=json
&fl=id
```

FIGURE 14 : LES PARAMETRES QUE L'API R&R A BESOIN POUR CREER LES CARACTERISTIQUES

   - q= la question
   - gt=l'id de la réponse, la note de pertinence
   - *generateHeader* : génération du *header* pour le fichier CSV training data
   - *rows* : les nombre de vecteurs de caractéristiques retournées
   - returnRSInput : un booléen qui indique si nous désirons recevoir les vecteurs de caractéristiques.
   - wt : le type de formatage de la réponse (JSON…)
2. Le service R&R transforme chaque couple question/réponse_candidate en un vecteur de caractéristiques suivie de la note de pertinence fournie dans la phase 1.

```
SInput=
cf818f7-9cc7-4c9c-b9dd-504399b8e82e,1.4919524,0.6315585,0.61277574,0.6315585,0.0,0.0,0.0,0.0,0.6666667,0,0.6931471805599453,2.0,2
cf818f7-9cc7-4c9c-b9dd-504399b8e82e,1.5216987,0.52356666,0.2634297,0.5861487,0.0,0.0,0.0,0.0,0.8333334,1,0.4054651081081644,1.0,4
```

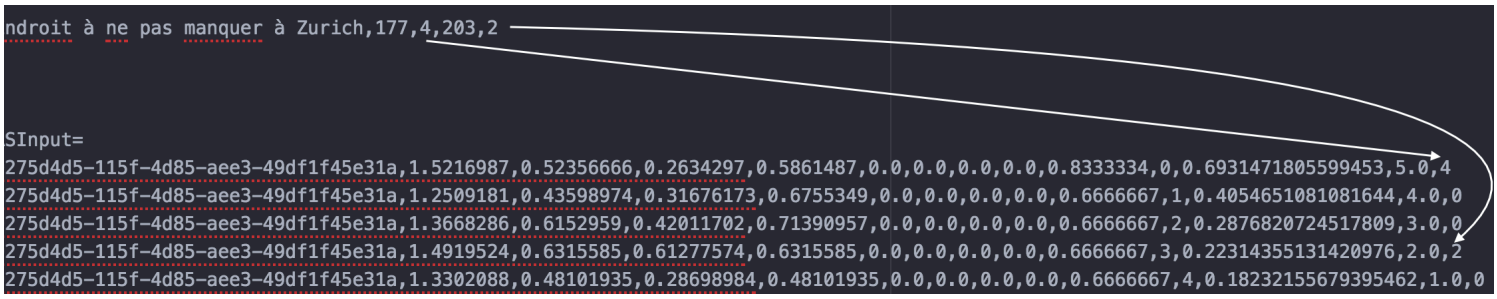
FIGURE 15: STRUCTURE DU RSIINPUT CONTENANT LES VECTEURS DE CARACTERISTIQUES

Voici la description des trois parties encadrées dans la figure ci-dessus :

- Partie 1 : associe une question avec une réponse que nous avons fourni au service dans la phase 1,
- Partie 2: représentent les caractéristiques utilisées par le service pour donner un score à la liaison entre la question et la réponse_candidates,
- Partie 3 : représente la note de pertinence.

A noter que si le nombre des pairs réponse_candidate/note de pertinence fourni dans le paramètre « gt » est plus petit que le nombre de lignes(rows) demandé, le service R&R donne une note de pertinence 0 aux autres réponses retournées par la requête.

Par exemple si nous soumettons une requête à l'API avec une question et 2 réponses candidates, alors que le nombre de rows demandé est 5, le service nous retourne les vecteurs de caractéristiques suivants :



'IGURE 16: TRANSFORMATION DE LA QUESTION EN VECTEURS DE CARACTERISTIQUES

Sur cette figure nous remarquons que l'API a retourné 5 vecteurs de caractéristiques, mais étant donné que nous avons fourni des notes de pertinence pour seulement 2 réponses candidates, il a donné automatiquement 0 aux autres vecteurs retournée par le Retriever.

3. Les phases 1 et 2 doivent être effectuées pour chaque question avec laquelle nous souhaitons entrainer le service R&R.
   Les vecteurs retournés par l'API sont rassemblés dans un seul fichier CSV. Ce fichier CSV (appelé le training data) doit être structuré de la manière suivante avant d'être envoyé à l'API pour créer et entrainer le *Ranker* :

```
query_id, feature1, feature2, feature3,...,ground_truth
question_id_1, 0.0, 3.4, -900,...,0
question_id_1, 0.5, -70, 0,...,1
question_id_1, 0.0, -100, 20,...,3
...
```

FIGURE 17: STRUCTURE DU TRAINING DATA

Voici un schéma[30] qui résume les étapes du Ranking dans le service R&R :

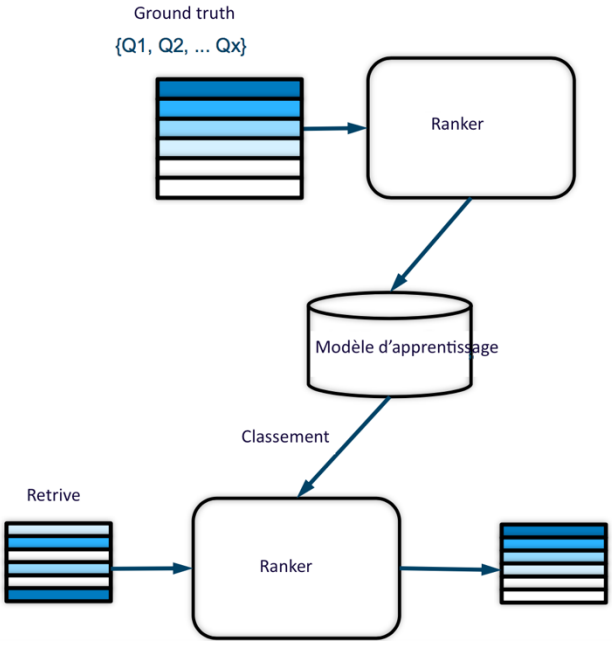


FIGURE 18: SCHEMA RESUMANT LES ETAPES DU RANKING

[30] https://www.slideshare.net/komine/watson-api-20160716-rev02

B) ENTRAINER LE RANKER EN UTILISANT LE SCRIPT TRAIN.PY

Pour simplifier la procédure du training, IBM a fourni un script (écrit en Python) qui fonctionne de la façon suivante :

1. Nous créons un fichier CSV où dans chaque ligne nous avons une question suivie des réponses possibles avec une note de pertinence(label) :

```
"my first query", "doc_id_1", "1", "doc_id_24", "3", "doc_id_7000", "1"
"my second query", "doc_id_36", "1", "doc_id_2", "3", "doc_id_3", "1"
```

FIGURE 19: STRUCTURE DU FICHIER CSV POUR LE SCRIPT TRAIN.PY

2. Par la suite le script créé automatiquement un fichier traning data(dans le « working directory ») contenant toutes les vecteurs de caractéristiques, et l'envoie à l'API R&R pour créer le Ranker et l'entrainer.

La figure suivante montre la façon d'utiliser le script train.py :

Running the script

The `train.py` Python script takes the following arguments:

```
-i {relevance_file} -c {cluster_id} -x {collection_name} [-r {solr_rows_per_query}] [-n {ranker_name}]
```

The `-r` and `-n` arguments are optional.

- Replace `{username}` and `{password}` with your service credentials.
- Replace `{relevance_file}` with the location of your relevance file.
- Replace `{cluster_id}` and `{collection_name}` with your information.
- **(Optional)** The `-r` argument specifies the number of answer results that the query returns. The correct answer must be in the results. The default value is 10.
- **(Optional)** Replace `{ranker_name}` with an name for the ranker that means something to you.

> **Note:** The `train.py` script requires Python major version 2 and fails if it is run under Python major version 3. To check the version of Python installed on your system, run the following command:

```
python --version
Python 2.7.10
```

FIGURE 20: MODE D'EMPLOI DU SCRIPT TRAIN.PY

A la fin nous récupérons l'*id* du *ranker* qui pourra être réutilisé pour ré-entrainer notre Ranker dans le futur.

Les documents CSV suivants, que nous allons annexer à ce travail, contiennent les données avec lesquels nous avons entrainé nos 4 instances de Ranker :

- tr4.csv contient 50 questions
- tr5.csv contient 100 questions
- tr6.csv contient 100 questions
- tr7.csv contient 100 questions (Les mêmes 50 en 2 fois)

Dans le chapitre Evaluation, nous allons donner plus de détails sur chacun de ces fichiers d'entraînement.

2. RE-ENTRAINER LE SYSTEME

Afin d'augmenter la pertinence des réponses du R&R, on peut remettre à jour le « ground truth » et répéter la phase d'entraînement. Voici un schéma[31] qui résume la structure du service R&R :

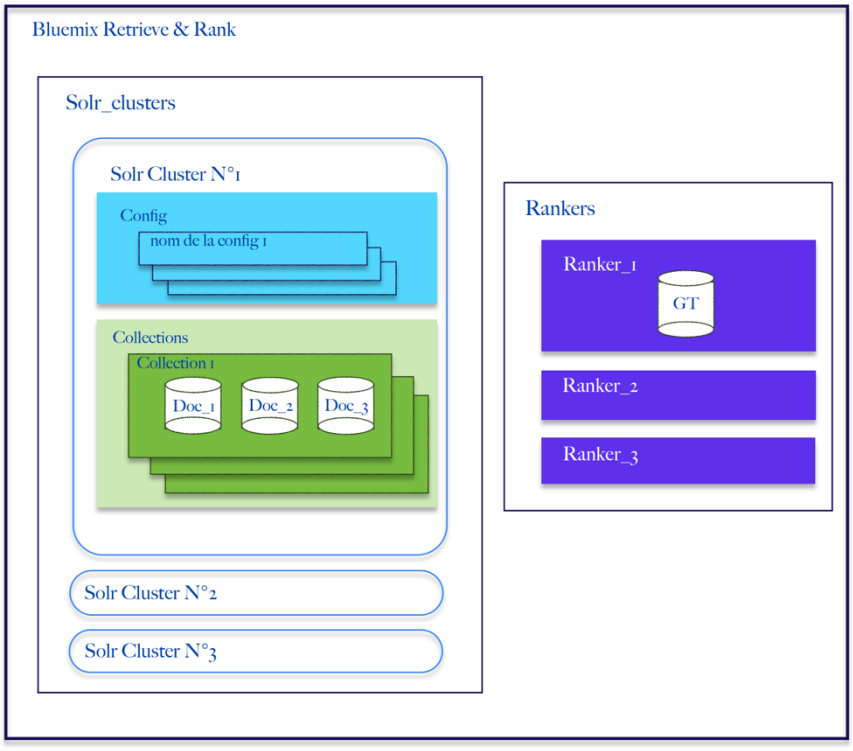


FIGURE 21: ARCHITECTURE DU SERVICE RETRIEVE AND RANK

[31] http://qiita.com/VegaSato/items/6d2d03d6a8b42adcf87e

IV. EVALUATION DES RESULTATS

Ces graphiques montrent l'évolution des réponses du service R&R par rapport au nombre et au types de questions que nous l'entrainons avec.

A. DONNÉES DU TRAINING

Nous avons évalué la pertinence des réponses du service R&R avec 50, 75, 100 questions.

Nous avons également testé le système avec un ficher nommé 50_50:

En effet, nous avons évalué la pertinence du système en lui soumettant 50 questions dans le « ground truth », ensuite nous lui avons posé les mêmes questions avec la note de la pertinence des trois premières réponses retournées et les avons mis dans un fichier csv pour pouvoir les utiliser dans le « ground truth » :

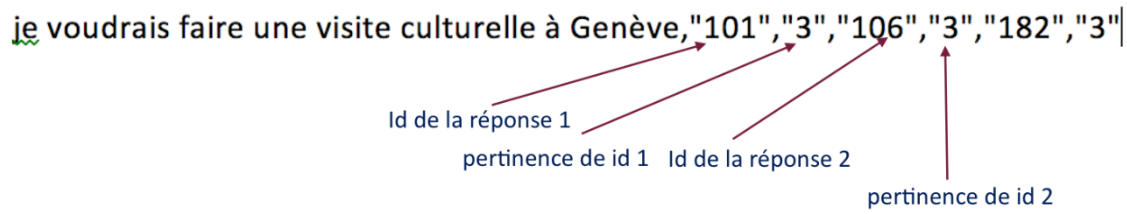


FIGURE 22: UNE LIGNE REPRESENTANT LES TROIS PREMIERES REPONSES DU SYSTEME A UNE REQUETE AVEC LEURS NOTES DE PERTINENCE

En effet cette expérience a été monter pour étudier l'importance feed-back que nous pouvons soumettre au service R&R sur ses propres réponses.

Pour résumé voici les étapes des la création des données d'entraînement 50_50 :

1. Entrainer le système avec 50 questions,
2. Envoyer des requêtes avec ces **même** 50 questions,
3. Noter la pertinence des 3 premières réponses sur chaque question et mettre dans un fichier csv avec le bon format du « ground truth »,
4. Soumettre au ranker ce fichier « ground truth » qui contiendra les 50 premières questions(étape 1) avec les autres 50 (créées lors de l'étape 3).

B. PERTINENCE

Nous avons donné une note de pertinence de la la réponse de la façon suivante :

- 4 une réponse parfaite,
- 3 réponses convenable mais pas ce que nous attendions, par exemple pas dans le bon endroit :
  - Nous demandons un restaurant italien mais nous donne un restaurant chinois,
  - Nous demandons un musée d'histoire mais la réponse est un musée d'art.
- 2 la réponse traite le sujet de la requête, par exemple :
  - La requête concerne un dîner et la réponse est endroit pour boire un café ou prendre le petit déjeuné.
- 1 pas de lien.

C. TYPES DE QUESTIONS DES REQUÊTES :

Nous avons testé le système de 2 façons :

- 50 questions que le système a utilisé pour l'entraînement,
- 50 nouvelles questions.

Nous avons fait en sorte que les questions soient parfois dans un langage formel et parfois dans un langage simple.

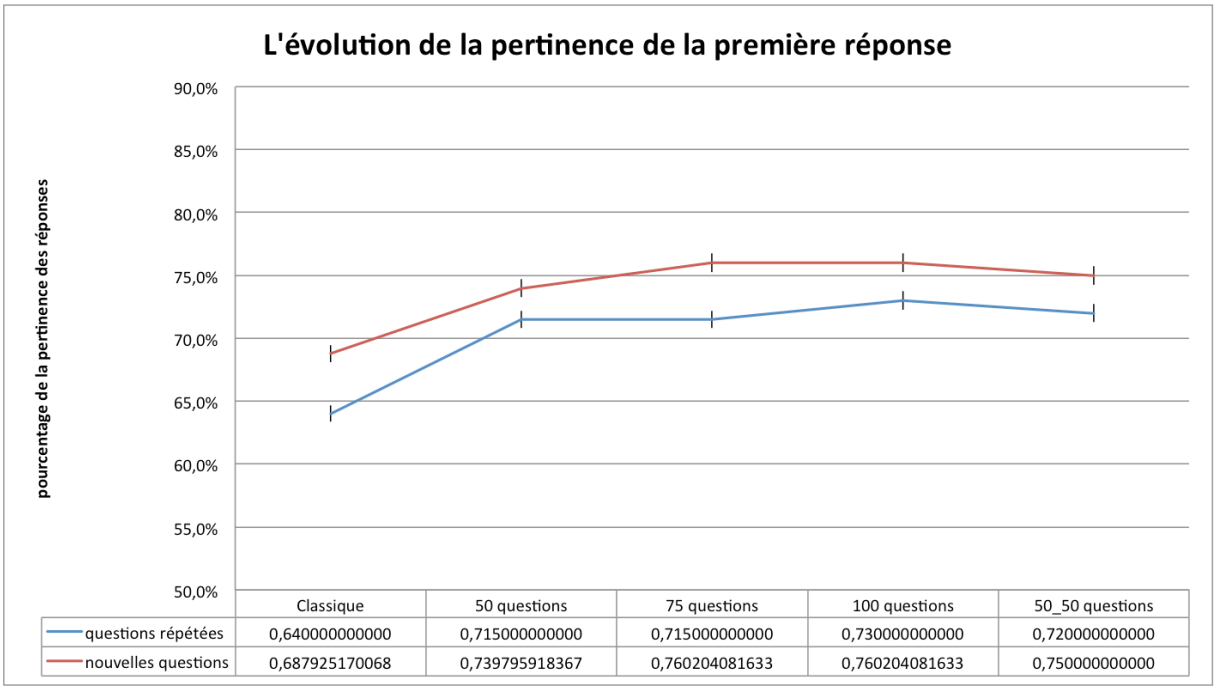
Dans le document annexé (annexe 7), vous trouverez les tests sur les différentes instances des Rankers, sur lesquels nous nous sommes basés pour créer les graphiques que vous allez trouver plus bas. Chaque ligne est représentée par la figure 22 (plus haut).

D. EVALUATION DES RÉSULTATS

Pour pouvoir comparer la pertinence des réponses du système, nous avons créé 2 graphiques.

1. L'EVOLUTION DE LA PREMIERE REPONSE :

Le premier permet de voir l'évolution de la pertinence de la première réponse entre les recherches classiques, et un système entrainé avec 50,75,100 et 50_50 questions:



	Classique	50 questions	75 questions	100 questions	50_50 questions
questions répétées	0,640000000000	0,715000000000	0,715000000000	0,730000000000	0,720000000000
nouvelles questions	0,687925170068	0,739795918367	0,760204081633	0,760204081633	0,750000000000

FIGURE 23: L'ÉVOLUTION DE LA PERTINENCE DE LA PREMIEEERE REPONSE

2. L'EVOLUTION DES TROIS PREMIERES REPONSES :

Ce deuxième graphique permet de voir l'évolution de la pertinence des 3 premières réponses entre les recherches classiques, et un système entrainé avec 50,75,100 et 50_50.

En sachant que le service R&R retourne les réponses classées par le degré de confiance, nous avons donné un poids différent à chacune des réponses :

- Un poids de 3 à la réponse que le service a classé en premier,
- Un poids de 2 à la réponse que le service a classé en deuxième,
- Un poids de 1 à la réponse que le service a classé en troisième,

Ce choix a été fait parce que le classement des réponses a une signification très importante pour le service R&R, en effet le service classe ses réponses par degré de confiance.

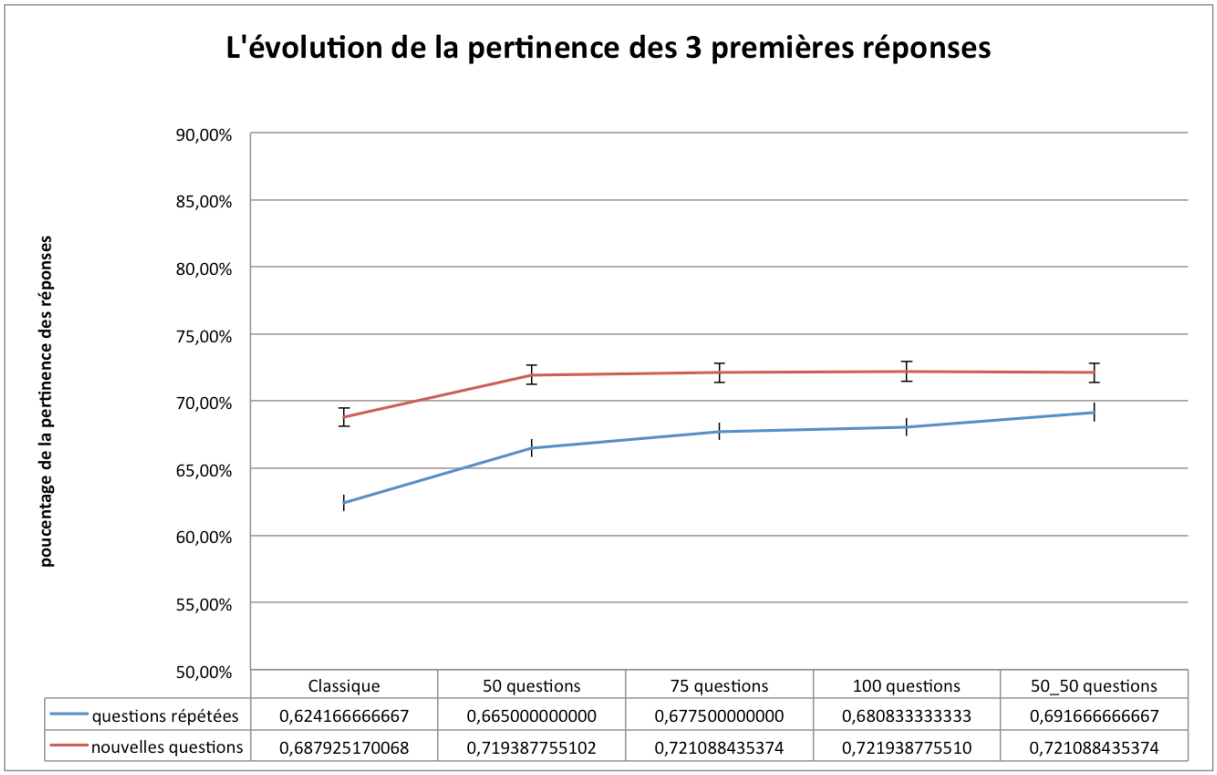


FIGURE 24: L'ÉVOLUTION DE LA PERTINENCE DES 3 PREMIERES RÉPONSES

3. A RETENIR

Ces graphiques nous montrent que les recherches du *Ranker* — qui utilise l'apprentissage d'ordonnancement sont beaucoup — plus pertinentes que celles des recherches classiques. Dans tous les cas au moins une augmentation de la pertinence de 5% a été remarquée avec les seules 50 premières questions.

Nous constatons que la pertinence des réponses du système augmente avec le nombre de questions avec lesquelles nous l'entraînons, bien que cette augmentation puisse paraitre parfois légère, comme c'est le cas pour les « nouvelles questions » (courbe rouge), mais la tendance reste positive.

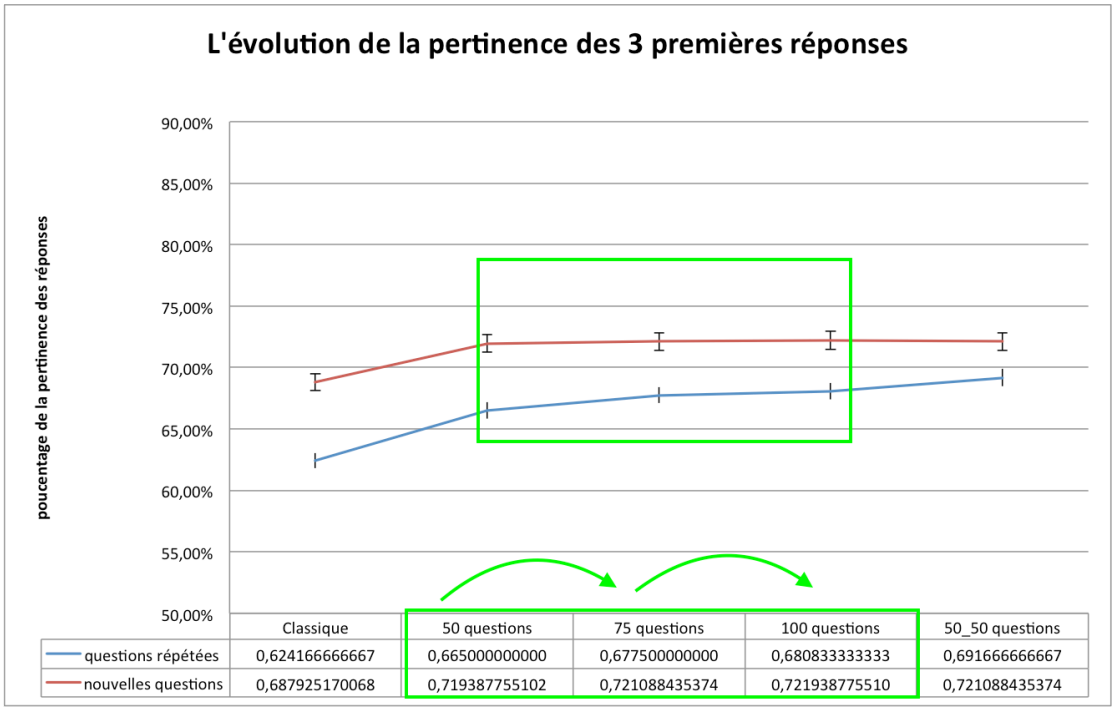
Cette augmentation est remarquée lorsque les questions font partie du fichier d'entraînement, mais également lorsque les questions sont nouvelles. Toutefois nous constatons que lorsque les questions font partie du fichier d'entraînement, la tendance positive de la courbe est plus importante. Ce dernier constat nous laisse penser que lorsqu'un type de question est très fréquent, le fait de retourner un *feedback* au service sur ses réponses peut augmenter la pertinence de ses réponses ultérieures.

Nous remarquons également que le type de questions peut avoir un impact sur la pertinence des documents retournés par le service R&R. Cela signifie que R&R doit être entrainé par des experts qui connaissent la nature des questions que l'application — qui utilise le service — a tendance à recevoir. Ceci confirme le fait que Watson est un « domain-oriented », car dans le développement d'une

application qui peut recevoir des questions génériques (comme ça été le cas pour le jeu télévisé Jeopardy !), l'entraînement peut s'avérer bien plus compliqué.

L'expérience du cas 50_50 qui retourne un *feedback* au service R&R indique que le système n'apprend pas les réponses mais extrait les caractéristiques les plus importantes qui l'aideront à répondre aux requêtes qu'il recevra dans le futur.

Nous avons remarqué également que l'évolution des performances a été plus forte lorsque nous étions passés de 50 à 75 questions (d'entraînement), que lorsque nous étions passés de 75 à 100.



	Classique	50 questions	75 questions	100 questions	50_50 questions
questions répétées	0,624166666667	0,665000000000	0,677500000000	0,680833333333	0,691666666667
nouvelles questions	0,687925170068	0,719387755102	0,721088435374	0,721938775510	0,721088435374

FIGURE 25: L'ÉVOLUTION DE LA PERTINENCE D'UN SYSTEME ENTRAÎNÉ AVEC 50, 75 ET 100 QUESTIONS

Pour en vérifier la raison, nous avons tenté dans un premier temps de savoir si le type dans les premières 25 questions était différent des 25 restantes avec lequel nous avons entraîné le système. Mais, nous n'avons pas trouvé que la nature de ces questions été très différentes.

Par contre nous avons remarqué que dans les restantes première 25, nous avions souvent mis dans le fichier d'entraînement une note de pertinence pour deux, trois ou quatre documents, en donnant aussi une note pour les documents moins pertinents (1,2 ou 3) :

```
où est ce que je peux profiter du beau temps à Genève,"18",4,"1","3"
visite pour les amateurs de l'art,"239",4
visite en couple à Geneve,"25","4","59",4
Je veux manger une pizza à Geneve,"74","4"
Manger des glaces à Genève,"75","4"
Boire un vers dans un endroit charmant,"71","4","75","4"
Je désire manger asiatique,"80","4","74","3"
```

Alors que pour les 25 questions restantes la note a été souvent donnée pour le document le plus pertinent :

```
Je veux voir les vieux cartiers de Bale,"266","4"
je veux boire un vers à Bale,"274","4"
boire un vers avec des amis,"279","4"
Visiter un musée à Zurich,"192","4"
Endroit à visiter à Zurich,"188","4"
Faire du sport à Zurich,"189","4"
```

En effet ceci voudrait dire que la *Ranker* arrive mieux à extraire les caractéristiques d'un couple requête/document lorsque nous lui indiquons des notes de pertinence et aussi de non pertinence sur plusieurs documents.

Parmi les problématiques qui peuvent détériorer les performances du service R&R, nous retrouvons le manque de plusieurs techniques de recherche d'information dans la partie Retrieve du service, — par exemple le fait qu'il ne prend pas en compte la racinisation en Français.

Comme nous l'avons mentionné précédemment, l'ordonnancement est effectué à partir des documents retournés par la recherche Solr. Or, la nature des documents sélectionnés peut ne pas inclure de mots clés, ce qui peut causer l'élimination des documents les plus pertinents dès la première phase(Retrieve).

Une solution pour être sûr d'avoir les bons résultats parmi les documents retournés par « Retrieve » est de modifier le paramètre « rows » (augmenter le chiffre), en sachant que, par défaut, il retourne 10 documents.

V. APPLICATION

L'objectif de cette application est de pouvoir soumettre les requêtes aux différentes instances du Ranker dont les étapes de création ont été mentionnées plus haut.

La figure 26 montre la fenêtre principale qui permet d'envoyer une requête en choisissant le mode de recherche :

- Recherche classique,
- Avec les différentes instances du Ranker (50, 75, 100, 50_50).

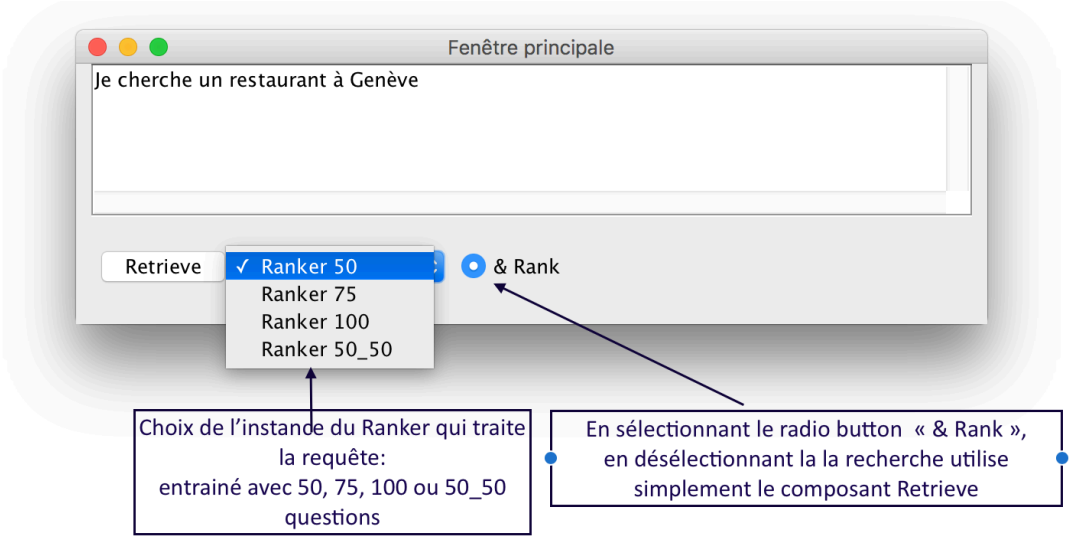


FIGURE 26: FENETRE PRINCIPALE DE L'APPLICATION

La figure 27 représente l'écran qui affiche les lieux retournés par le service ainsi que la description et score de confiance pour chaque lieu(après une requête).

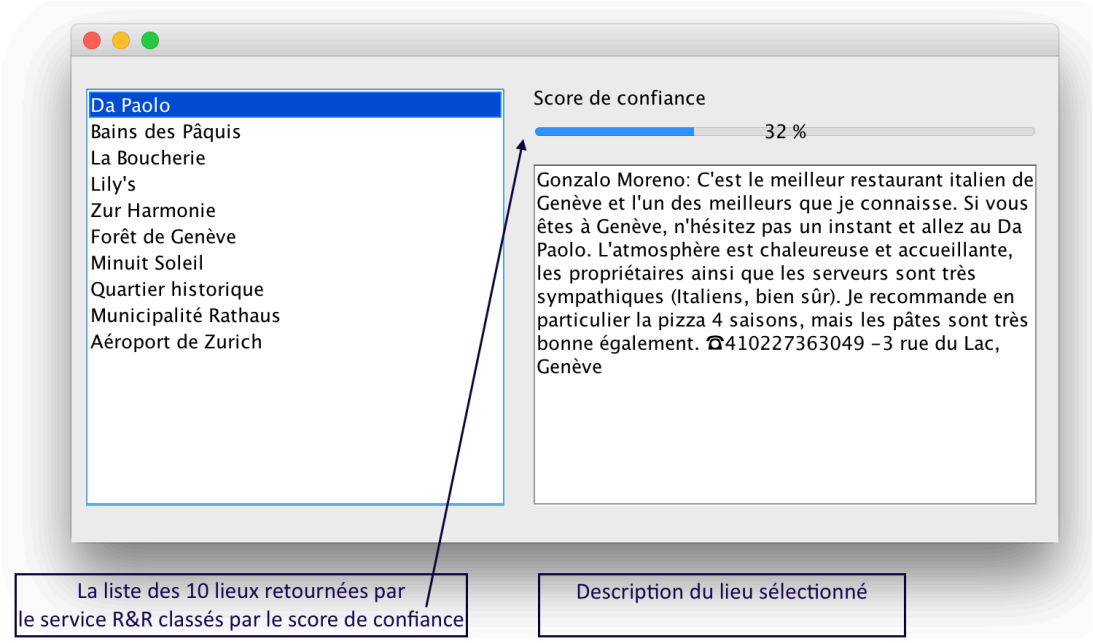


FIGURE 27: FENETRE DE LA LISTE DES LIEUX RETOURNES PAR LE SERVICE R&R

VI. CONCLUSION

A travers ce document, nous avons essayé d'exposer les différents composants du service Retrieve and Rank. Nous avons constaté que chacun de ses composants représente un élément important dont la maitrise influence fortement la pertinence des recherches.

En effet, à travers cette architecture décomposée, IBM tente de donner le moyen de mieux adapter les différents éléments du service aux objectifs visés, par exemple le fait que chaque composant donne un résultat partiel, nous donne la possibilité d'inclure dans le calcul du degré de confiance d'autres critères — comme les avis des clients —, ce qui améliorer la pertinence des documents retournés.

Ce service peut également être intégré dans une architecture beaucoup plus riche, avec d'autres services Watson comme le montre la figure 28.

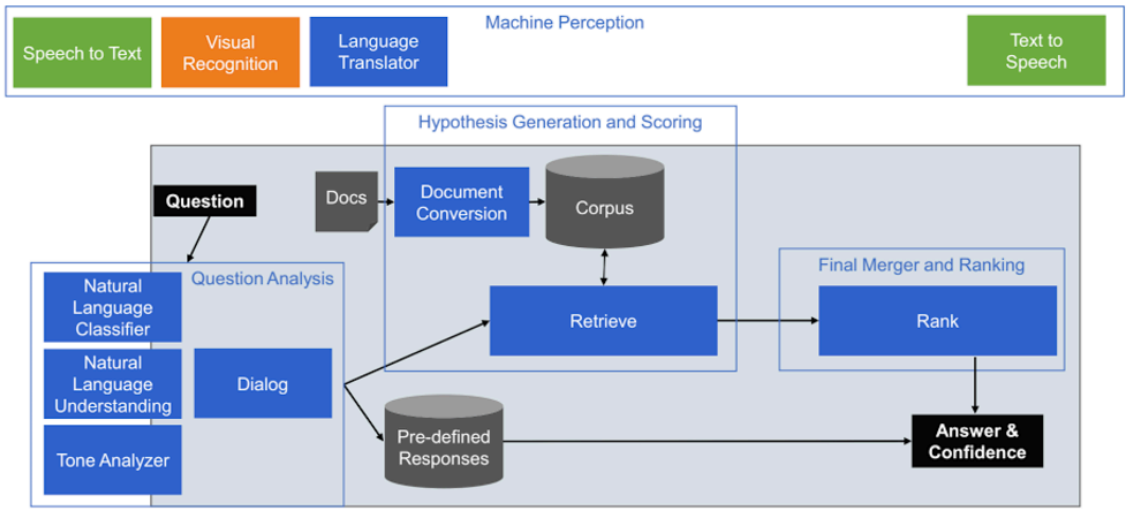


FIGURE 28: EXEMPLE D'INTÉGRATION DU SERVICE R&R AVEC D'AUTRES SERVICES[32]

Nous avons également pu constater que la maitrise de chaque étape est importante dans le processus de recherche, par exemple une mauvaise configuration de la partie Solr, peut influencer fortement les documents retournés par le composant Retrieve, ce qui aura comme conséquence de fournir au *Ranker* des documents non pertinents.

La connaissance du domaine lié à l'application à développer (Médecine, Sécurité informatique etc..) est également très importante, en effet seul les experts d'un

[32] Livre : *Building Cognitive Applications with IBM Watson Services : Volume 1 Getting started , IBM RedBook, 06.2017, Etas Unis*

domaine peuvent juger la pertinence des documents ou encore mettre dans le « ground Thruth » les questions qui ont le plus de chance d'être posées.

L'amélioration des résultats est corrélée à constater l'efficience de l'entraînement du système. Non seulement le nombre de questions d'entraînement dans le « ground truth » est important mais également la nature des questions, le cas 50_50 illustre bien le fait que les *feedbacks* que nous donnons au système peuvent fortement contribuer à améliorer ses performances pour un type spécifique de questions.

Ces améliorations qu'offre l'intelligence artificielle dans le domaine des recherches textuelles, peuvent être très bénéfiques pour les entreprises et les particuliers, et il serait dommage de ne pas en profiter.

BIBILIOGRAPHIE

[1]. Ensemble de documentation et APIs, et forum pour les services Watson, 2017, Par Ibm Watson disponibles à l'adresse : https://www.ibm.com/watson/developercloud/doc/index.html

[2]. Daisuke Sato, 03.12.2015, Traduit du Japonais , disponible sur le site : http://qiita.com/VegaSato/items/6d2d03d6a8b42adcf87e

[3]. Daisuke Sato, 03.12.2015, Traduit du Japonais , disponible sur le site : http://qiita.com/VegaSato/items/6b844bbd73a97ad58e37

[4]. Hiroaki Komine, Watson API トレーニング, 2016, traduit du Japonais, disponible sur le site https://www.slideshare.net/komine/watson-api-20160716-rev02

[5]. Pierre-Carl Langlais, Le deep learning est-il le futur du text mining ? 02,09,2016, disponible sur le site : https://scoms.hypotheses.org/657

[6]. IMEX research,2014, The new challenges of unstructured data, disponible sur le site : http://www.imexresearch.com/newsletters/obs.html

[7]. Journal of research and development, IBM, Vol. 56, No. 3/4, May/July 2012 , USA

[8]. Présentation de SolrCloud Solr(4.0), 30.05.2013, Dominique, disponible sur le site https://www.eolya.fr/2013/05/30/presentation-de-solrcloud-solr-4-0/

[9]. Nicolas Travers, Raphaël Fournier S'niehotta et Philippe Rigaux , 2014-2017, Disponible sur le site : http://b3d.bdpedia.fr/solr.html

[10]. Xavier Lecomte, La recherche Full Text avec Solr, Configurer un moteur de recherche performant à l'aide d'Apache Lucene/Solr et Apache Tomcat, disponible sur le site : http://g-rossolini.developpez.com/tutoriels/solr/?page=schema , DEVELOPPEZ LLC, Company Registration Number : 3734566

[11]. Nima Asadi・Jimmy Lin, Document vector representations for feature extraction in multi-stage document ranking, 20.05.2012, disponible sur le site : https://cs.uwaterloo.ca/~jimmylin/publications/Asadi_Lin_IRJ2013.pdf

[12]. Chris Ackerson, Developing with IBM Watson Retrieve and Rank, 2016, Article en trois parties, disponible sur le site : https://medium.com/machine-learning-with-ibm-watson/developing-with-ibm-watson-retrieve-and-rank-part-1-solr-configuration-29c18e52966f

[13]. Exposées dirigés par Dominique Revuz et Etienne Duris,Apache Solr,2008, exposées d'étudiants en troisième année d'école d'ingénieurs l'UFR ingénieurs 2000 de l'Université Paris-Est/Marne-la-vallé, en filière Informatique et Réseaux disponibles sur le site : http://igm.univ-mlv.fr/~dr/XPOSE2008/Apache%20Solr/index.html

[14]. Bob Dill IBM Distinguished Engineer and CTO, Zero to cognitive, chapter 11 Rank and retrieve, 02.2017, vidéo disponible sur internet : https://www.youtube.com/watch?v=L-82Yr5sDH8

[15]. Fartash Haghani, Create a Natural Language Question Answering system with IBM Watson, 14.04.2016, article disponible sur le site : http://fartashh.github.io/post/qa-system-watson/
[16]. Amol Sonawane ,Utiliser Apache Lucene pour effectuer des recherches textuelles, 18.08.2009, disponible sur le site : https://www.ibm.com/developerworks/java/library/os-apache-lucenesearch/index.html
[17]. Apprentissage profond, Article disponible sur Wikipédia : https://fr.wikipedia.org/wiki/Apprentissage_profond
[18]. Michael Nilsson & Diego Ceccarelli, Bloomberg LP, Learning to Rank in Solr: Presented, 2015,disponible sur le site : https://fr.slideshare.net/lucidworks/learning-to-rank-in-solr-presented-by-michael-nilsson-diego-ceccarelli-bloomberg-lp
[19]. Laporte, Léa and Déjean, Sebastien and Mothe, Josiane Sélection de variables en apprentissage d'ordonnancement : évaluation des SVM pondérés. (2015) Document numérique, vol. 18 (n° 1). pp. 97-121. ISSN 1279-5127
[20]. Laporte Léa, La sélection de variables en apprentissage d'ordonnancement pour la recherche d'information : vers une approche contextuelle, 18 novembre 2013, En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE, disponible sur le site : http://thesesups.ups-tlse.fr/2170/1/2013TOU30240.pdf
[21]. Learning to Rank, Article Wikipédia, Url : https://en.wikipedia.org/wiki/Learning_to_rank#Feature_vectors
[22]. Building Cognitive Applications with IBM Watson Service, 06.2017 : Volume 1 Getting started, IBM RedBook, disponible sur internet : https://books.google.co.ma/books?id=7W0pDwAAQBAJ&printsec=frontcover&dq=Building+Cognitive+Applications+with+IBM+Watson+Services:+Volume&hl=fr&sa=X&ved=0ahUKEwiO3JiaoILWAhVHSBQKHa_5AgYQ6AEIUjAF#v=onepage&q=Building%20Cognitive%20Applications%20with%20IBM%20Watson%20Services%3A%20Volume&f=false , Etas unis
[23]. Kshitij Fadnis, IBM Watson: DeepQA Framework and Retrieve and Rank Integration, 13.01.2016, video disponible sur internet : https://www.youtube.com/watch?v=pupGatYlfqQ
[24]. Code open source par IBM sur le lien : https://github.com/watson-developer-cloud/java-sdk
[25]. Naveen Bans, Soumyajit De, Sanober Nishat, Question / Answering system &Watson , 2012, https://www.cse.iitb.ac.in/~cs626/cs626-sem1-2012/seminar/grp3-seminar-watson.ppt.pptx

ANNEXE 1 : ÉTAPES POUR AVOIR LES CREDENTIALS D'UN SERVICE WATSON

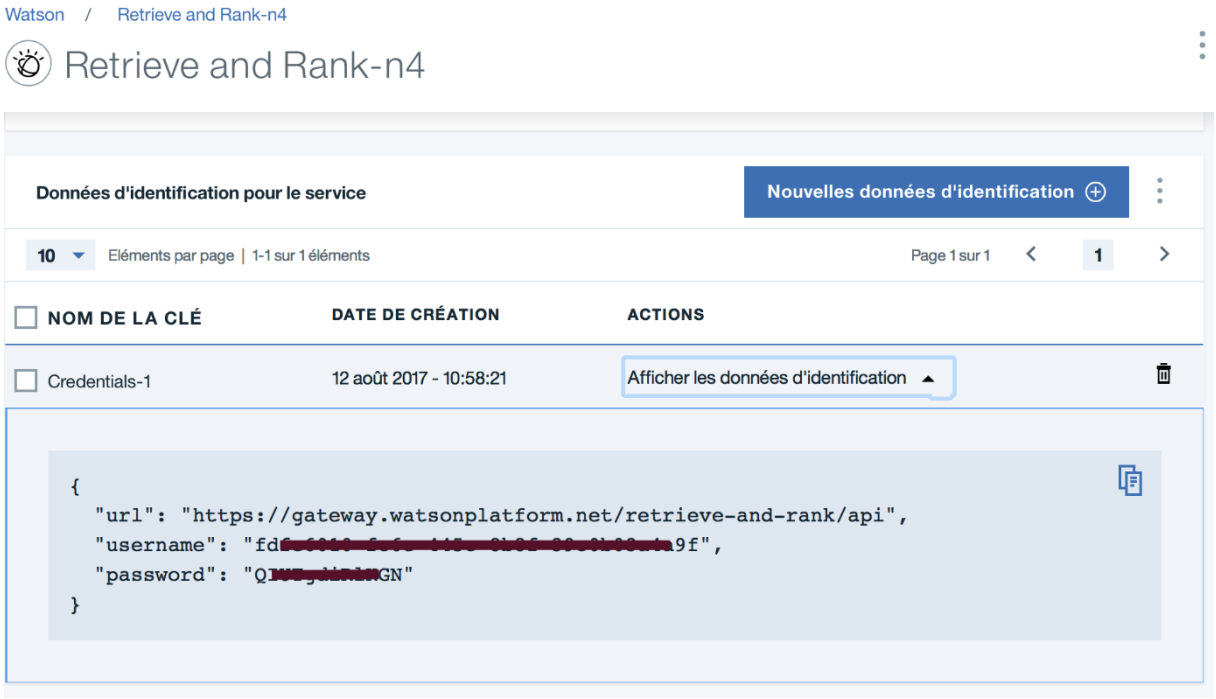
Stage 1: Get your service credentials

Before you can work with a service in Bluemix, you need service credentials. If you already have credentials for the Retrieve and Rank service, you can skip this stage.

To get your service credentials, follow these steps:

1. Log in to Bluemix.
2. Create an instance of the service:
   a. In the Bluemix **Catalog**, select the Retrieve and Rank service.
   b. Under **Add Service**, type a unique name for the service instance in the **Service name** field. For example, type `rr_tutorial_{username}`, and replace {username} with your name. Leave the default values for the other options.
   c. Click **Use**.
3. Copy your credentials:
   a. On the left side of the page, click **Service Credentials** to view your service credentials.
   b. Copy `username` and `password` from these service credentials. You'll need them in the following stages.

Pour chaque service nous aurons les *credentials* comme le montre la capture d'écran suivante :



ANNEXE 2 : PARAMETRAGE DES DOCUMENTS EN ENTREE DANS LE SERVICE DOCUEMENT CONVERSION

Key	Description
`level`	Required when defining any other font value. Heading level to be generated. When specifying a list of heading values, the first configuration in each heading list that matches something from the input document is the configuration that is applied to the conversion to create that heading type. Valid values are `1` (h1), `2` (h2), `3` (h3), `4` (h4), `5` (h5), and `6` (h6).
`min_size`	Minimum size font to be converted into the defined heading level. Default is `0`. Valid values are any font size above `0`.
`max_size`	Maximum size font to be converted into the defined heading level. Valid values are any font size larger than the minimum font size.
`bold`	Matches if the input is bold. Set to `true` when the input font that you want to be converted into a particular heading is bold. Set to `false` when the font isn't bold. Not setting this configuration means that the conversion ignores whether the input is bold.
`italic`	Matches if the input is italic. Set to `true` when the input font that you want to be converted into a particular heading is italic. Set to `false` when the font isn't italic. Default value is `false`.
`name`	The name of the font. If not defined all font names will be matched. Valid values are font names.

Table 1. Microsoft Word input heading font configurations

Key	Description
`level`	Required when defining any other style value. Heading level to be generated. When specifying a list of heading values, the first configuration in each heading list that matches something from the input document is the configuration that is applied to the conversion to create that heading type Valid values are `1` (h1), `2` (h2), `3` (h3), `4` (h4), `5` (h5), and `6` (h6).
`names`	A list of style names that should be considered a heading in the conversion. Valid values are an array of style names.

Table 2. Microsoft Word input heading style configurations

ANNEXE 3 : LES OPTIONS POUR PASSER D'UN FICHIER HTML À UN FICHIER HTML NORMALISE

Key	Description
`exclude_tags_completely`	Defines tags that should be removed completely with their content. Valid values are a list of tags.
`exclude_tags_keep_content`	Defines tags that should be removed, but content is kept. Valid values are a list of tags.
`keep_content.xpaths`	Specifies a list of XPaths that identify content that is converted. If this value is set, anything that matches one of the XPaths are included in the output. The inclusions specified by this parameter are processed after any processing specified by `exclude_content_xpaths`.
`exclude_content.xpaths`	Specifies the values to identify the main content that is not converted. If this value is set, anything that matches one of the XPaths is excluded from the output.
`keep_tag_attributes`	List of attributes to keep in the HTML tags. Can only be specified if exclude_tag_attributes is not specified. Valid values are `EVENT_ACTIONS`, , or an array of individual attributes. Selecting `EVENT_ACTIONS` includes all of the JavaScript action attributes. Selecting includes all attributes. For best results, include `rowspan`, `colspan`, and `border`.
`exclude_tag_attributes`	List of attributes to strip from the HTML tags. Can only be specified if keep_tag_attributes is not specified. Valid values are `EVENT_ACTIONS`, , or an array of individual attributes. Selecting `EVENT_ACTIONS` excludes all of the JavaScript action attributes. Selecting excludes all attributes. For best results, do not exclude `rowspan`, `colspan`, or `border`.

ANNEXE 4 : CODE DE LA CRÉATIONDOCUMENT CONVERSION ET INDEXATION DANS R&R

Partie 1

```
final String versionDate = "2017-08-01";
//  Création d'instance Document conversion
DocumentConversion serviceDC = new DocumentConversion(versionDate);

serviceDC.setUsernameAndPassword(USERNAMEDC, PASSWORDDC);

//  le document word où se trouvent les lieux touristiques
final File word = new File("/Users/basshass/Desktop/RAR/memoir/guideMain.docx");

JsonParser jsonParser = new JsonParser();

// La configuration pour le service document conversion
String configAsString2 = "{"
      + "\"answer_units\":{"
      + "\"heading\":{"
      + "\"fonts\":["
      + "{\"level\":2,\"min_size\":1}"
      + "]}}}";

//   Transformation de la config d'un String (en Format JSON) en un Object Json
JsonObject customConfig = jsonParser.parse(configAsString2).getAsJsonObject();

System.out.println("conversion et découpage du docuemnt word en Answers Units ");

final Answers wordToAnswersWithCustomConfig =
    serviceDC.convertDocumentToAnswer(word, null, customConfig).execute();
```

urtie 2

```
int cptId=1;
while (itAnswersUnits.hasNext()) {
    AnswerUnits ans = (AnswerUnits) itAnswersUnits.next();
    System.out.println("Indexing document...");
    // à noter que ces fileds son ceux qu'on créé dans le schema.xml de la config
    SolrInputDocument document1 = new SolrInputDocument();

    // Ici je filtre les ANswers Units en ignorant(ne pas ajouter à Solr) ceux où le texte vide
    if(ans.getContent().get(0).getText().length()!=0){
        document1.addField("id", cptId);
        document1.addField("title", ans.getTitle());
        document1.addField("content_text", ans.getContent().get(0).getText());
        listSolrInputDocuments.add(document1);

    }
    cptId++;

}
// je rajoute la liste des "SolrInputDocument" dans la collection Solr créée précédement

final UpdateResponse addResponse1 = solrClient.add("collectionb", listSolrInputDocuments);
System.out.println(addResponse1);

// je commit pour confirmer(si non il ne l'enregistre pas)
solrClient.commit("collectionb");
System.out.println("Indexed and committed document.");
// les documents
service = new RetrieveAndRank();
service.setUsernameAndPassword(USERNAMERR, PASSWORDRR);
solrClient = getSolrClient(service.getSolrUrl(SOLR_CLUSTER_ID), USERNAMERR, PASSWORDRR);

// transformation des answers_units recus de document conversion en liste Iterator
// sur cette liste on peut filtrer (par exp: supprimer ceux qui on le titre en format chiffre)
// pour le tranmettre ensuite à Solr

//   avoir la liste des Answers Units
List<AnswerUnits> answersunits =wordToAnswersWithCustomConfig.getAnswerUnits();

//Je crée une collection de SolrInputDocument
Collection<SolrInputDocument> listSolrInputDocuments =new ArrayList();

// Itérer sur le liste des Answers Units
Iterator<AnswerUnits> itAnswersUnits=answersunits.iterator();
```

ANNEXE 5 : CODE DE LA CRÉATION D'UN CLUSTER SOLR

```
public class CreateASolrClusterExample {
    private static final String USERNAMERR = "f[redacted]4a9f";
    private static final String PASSWORDRR = "([redacted])";
  public static void main(String[] args) throws InterruptedException {

    // 1 Création d'une instance de notre service R&R avec nos credentials ( voir annexe 1)
    RetrieveAndRank service = new RetrieveAndRank();
    service.setUsernameAndPassword(USERNAMERR, PASSWORDRR);

    // 2 Création du cluster Solr
    SolrClusterOptions options = new SolrClusterOptions("<cluster-name>", 1);
    SolrCluster cluster = service.createSolrCluster(options).execute();
    System.out.println("SolrCluster: " + cluster);

    // Attendre que le cluster Solr soit disponible
    // chaque 10 secondes il nous affiche l'état du cluster
    while (cluster.getStatus() == Status.NOT_AVAILABLE) {
      Thread.sleep(10000); // sleep 10 seconds
      cluster = service.getSolrCluster(cluster.getId()).execute();
      System.out.println("SolrCluster status: " + cluster.getStatus());
    }

    // 3 list Solr Clusters pour récupérer le cluster_id
    System.out.println("Solr clusters: " + service.getSolrClusters().execute());

  }
```

ANNEXE 6 : CODE DE CRÉATION D'UNE COLLECTION SOLR

```
// Création d'une instance de notre service R&R
service = new RetrieveAndRank();
service.setUsernameAndPassword(USERNAMERR,PASSWORDRR);
//Récupérer le cluster créé précédement avec Cluster_ID (annexe 5)
ServiceCall<SolrCluster> cluster = service.getSolrCluster(SOLR_CLUSTER_ID);
System.out.println(cluster.execute());
//création de la configuration
//avec le path où se trouve la configuration
File configZip = new File("/Users/basshass/conf2.zip");

// ajouter la configuration au Cluster
service.uploadSolrClusterConfigurationZip(SOLR_CLUSTER_ID,"configfrb", configZip).execute();

//création de la collection
CollectionAdminRequest.Create createCollectionRequest =new CollectionAdminRequest.Create();
// lui donner un nom
createCollectionRequest.setCollectionName("collectionb");

// Lier la configuration avec la collection
createCollectionRequest.setConfigName("configfrb");

// Créer une instance de SolrClient pour lancer la requête de création de la Collection à Solr
SolrClient solrClient= getSolrClient(service.getSolrUrl(SOLR_CLUSTER_ID), USERNAMERR,PASSWORDRR);
System.out.println("Creating collection...");
CollectionAdminResponse response = createCollectionRequest.process(solrClient);
    if (!response.isSuccess()) {
      System.out.println(response.getErrorMessages());
      throw new IllegalStateException("Failed to create collection: "
          + response.getErrorMessages().toString());
    }
System.out.println("Collection created.");
```

ANNEXE 7 : JEUX DE TESTES

1. QUESTIONS REPETEES

Recherche classiques[33] **:**

je veux manger à Lausanne 4,3,1

j'aimerais faire une visite aux sites historiques de Geneve 1,1,4

je veux visiter les magasins de Genève 2,3,4

j'aimerais visiter Genève pour voir les boutiques de luxe 4,4,1

J'aimerais faire une visite familiale à Lausanne,2,2,2

je voudrais faire une visite culturelle à Genève 1,3,3

je cherche un restaurant à Lausanne 1,4,2

que puis je visiter à Genève avec ma famille 4,4,1

je voudrais visiter un musée d'art sur Genève 4,4,3

je veux me balader à pied à Genève 3,1,1

où est ce que je peux voir les paysages de Genève1,1,3

je veux prendre le bateau a Geneve1,1,1

visiter avec les enfants a geneve4,2,1

j'aimerais me relaxer à Geneve1,1,1

je peux voir quoi à Genève 1,1,1

je voudrais me promener dans les lieux culturelles de Genève1,3,1

excursions à Geneve1,1,1

Auriez vous quelques conseils de places à visiter dans un rayon raisonnable autour de Bern1,1,1

je veux voir un musée typique suisse à Genève 2,1,2

Visite pour les curieux à Geneve 2,1,2

[33] *NB : Dans cette recherche nous n'avons pas mis l'identifiant des documents avant la note de pertinence.*

Les jolies paysages de Genève 3,3,1

Monuments de Genève 3,3,3

Visite religieuse 3,3,1

Spectacle à Genève 4,4,1

Belle vue sur Genève 4,3,3

Les adeptes de la science 1,4,4

Amoureux de la science 4,4,1

Visiter Geneve pour les amoureux de la nature 3,1,1

Symbole fort de Genève 3,4,1

Les amoureux de la litérature 2,1,2

Un chemin de promenade pour les passionnés de la nature 4,1,4

Un chemin de promenade pour les passionnés d'histoire1,4,4

randonnées à Geneve 1,1,3

Une balade à travers des années d'histoire 4,1,3

Shopping Genève 1,3,4

Joli pont à Genève 2,3,3

Place très celebre à Genève 4,4,1

Amoureux de la littérature 1,3,3

Maisons célèbres 4,4,4

Jardin au centre de Genève 3,4,3

Panorama de Genève 4,3,4

Visite guidée Genève 4,1,2

Endroit historique au centre de Genève 4,2,4

Statue à Genève 4,3,3

Acheter du fromage suisse 4,4,1

déguster du fromage Suisse 4,4,2

Je veux manger au centre ville de Genève 2,3,4

Je veux manger traditionnel suisse 4,2,3

Je voudrais manger typique suisse à Genève 1,3,3

Je veux de la bouffe italienne à Genève 1,1,1

Recherche entrainé 50 questions:

je veux manger à Lausanne,17,3,121,4,160,3

j'aimerais faire une visite aux sites historiques de Geneve,166,3,1,1,38,1

je veux visiter les magasins de Genève,254,1,15,4,243,3

j'aimerais visiter Genève pour voir les boutiques de luxe,144,2,15,4,171,3

J'aimerais faire une visite familiale à Lausanne,17,4,227,2,24,3

je voudrais faire une visite culturelle à Genève,106,2,11,4,101,3

je cherche un restaurant à Lausanne,123,4,8,1,276,3

que puis je visiter à Genève avec ma famille 59, 4,238,3,24,4

je voudrais visiter un musée d'art sur Genève 34,4,45,4,238,3

je veux me balader à pied à Genève 258,3,41,4,87,1

où est ce que je peux voir les paysages de Genève,18,4,246,1,254,1

je veux prendre le bateau a Geneve,1,1,18,1,238,2

visiter avec les enfants a geneve,38,4,78,2,92,1

j'aimerais me relaxer à Geneve,228,1,87,1,1,1

je peux voir quoi à Genève,238,2,246,1,254,2

je voudrais me promener dans les lieux culturelles de Genève,65,4,17,3,87,1

excursions à Geneve,224,3,1,1,231,3

Auriez vous quelques conseils de places à visiter dans un rayon raisonnable autour de Bern,238,4,184,2,89,1

je veux voir un musée typique suisse à Genève,73,1,7,3,1,45,3

Visite pour les curieux à Geneve,38,1,235,3,89,1

Les jolies paysages de Genève,143,3,69,1,203,3

Monuments de Genève,64,1,66,3,56,3

Visite religieuse 66,4,92,3,249,4

Spectacle à Genève,64,1,126,1,59,1

Belle vue sur Genève,59,4,137,3,66,4

Les adeptes de la science 36,4,62,4,79,1

Amoureux de la science,36,4,62,4,8,1

Visiter Geneve pour les amoureux de la nature,190,4,36,1,38,2

Symbole fort de Genève,31,4,130,3,155,1

Les amoureux de la litérature,190,2,36,2,124,2

Un chemin de promenade pour les passionnés de la nature,114,2,163,4,4,4

Un chemin de promenade pour les passionnés d'histoire,237,1,162,2,34,4

randonnées à Geneve,224,4,1,1,142,3

Une balade à travers des années d'histoire,238,2,246,4,272,4

Shopping Genève,59,1,64,4,66,1

Joli pont à Genève,56,2,40,3,48,3

Place très celebre à Genève,2,2,109,1,50,4

Amoureux de la littérature,67,3,54,3,102,2

Maisons célèbres 251,4,183,1,208,1

Jardin au centre de Genève,69,1,37,3,12,1

Panorama de Genève,64,1,66,4,56,4

Visite guidée Genève,66,4,47,2,19,4

Endroit historique au centre de Genève,69,4,173,3,56,4

Statue à Genève,58,4,64,1,126,1

Acheter du fromage suisse,73,4,260,2,82,4

déguster du fromage Suisse,73,4,226,4,76,2

Je veux manger au centre ville de Genève,70,4,12,2,77,4

Je veux manger traditionnel suisse,217,3,79,3,80,3

Je voudrais manger typique suisse à Genève,73,4,217,3,80,3

Je veux de la bouffe italienne à Genève,84,3,2,1,87,2

Recherche entrainé avec 75 questions:

je veux manger à Lausanne,121,4,160,3,217,3

j'aimerais faire une visite aux sites historiques de Geneve,166,3,1,1,38,1

je veux visiter les magasins de Genève,254,1,243,3,15,4

j'aimerais visiter Genève pour voir les boutiques de luxe, 171,3, 15,4,144,2

J'aimerais faire une visite familiale à Lausanne,17,4,227,2,24,3

je voudrais faire une visite culturelle à Genève,106,2,11,4,244,4

je cherche un restaurant à Lausanne,123,4,276,3,18,2

que puis je visiter à Genève avec ma famille,59,4,24,4,238,3

je voudrais visiter un musée d'art sur Genève,34,4,45,4,94,3

je veux me balader à pied à Genève,258,3,87,1,41,4

où est ce que je peux voir les paysages de Genève,18,4,213,1,254,2

je veux prendre le bateau a Geneve,18,1,1,1,238,2

visiter avec les enfants a geneve,78,2,38,4,92,1

j'aimerais me relaxer à Geneve,228,1,1,1,87,1

je peux voir quoi à Genève,238,2,246,1,254,2

je voudrais me promener dans les lieux culturelles de Genève,65,4,17,3,87,1

excursions à Geneve,224,3,1,1,231,3

Auriez vous quelques conseils de places à visiter dans un rayon raisonnable autour de Bern,238,4,184,2,46,1

je veux voir un musée typique suisse à Genève,73,1,216,1,45,3

Visite pour les curieux à Geneve,38,1,89,1,1,2

Les jolies paysages de Genève,37,4,143,3,69,1

Monuments de Genève,64,1,56,3,66,3

Visite religieuse,66,4,92,3,249,4

Spectacle à Genève,64,1,126,1,56,1

Belle vue sur Genève,59,4,137,3,66,4

Les adeptes de la science,36,4,62,4,127,1

Amoureux de la science,36,4,62,4,8,1

Visiter Geneve pour les amoureux de la nature,190,4,36,1,38,2

Symbole fort de Genève,31,4,130,3,155,1

Les amoureux de la litérature,190,2,36,2,124,2

Un chemin de promenade pour les passionnés de la nature,114,2,136,2,163,4

Un chemin de promenade pour les passionnés d'histoire,237,1,67,4,125,4

randonnées à Geneve,224,4,1,1,142,3

Une balade à travers des années d'histoire,238,2,246,4,272,4

Shopping Genève,64,4,56,1,59,1

Joli pont à Genève,56,2,40,3,167,3

Place très celebre à Genève,2,2,69,3,66,3

Amoureux de la littérature,67,3,54,3,117,2

Maisons célèbres,251,4,183,1,208,1

Jardin au centre de Genève,69,1,12,1,5,1

Panorama de Genève,64,1,56,4,66,4

Visite guidée Genève,47,2,19,4,66,4

Endroit historique au centre de Genève,69,4,56,4,173,3

Statue à Genève,58,4,64,1,56,2

Acheter du fromage suisse,73,4,82,4,226,4

déguster du fromage Suisse,73,4,226,4,210,4

Je veux manger au centre ville de Genève,70,4,79,4,77,4

Je veux manger traditionnel suisse,217,3,79,3,80,3

Je voudrais manger typique suisse à Genève,73,4,217,3,80,3

Je veux de la bouffe italienne à Genève,2,1,84,3,87,2

Recherche entrainé avec 100 questions :

je veux manger à Lausanne,121,4,160,3,217,3

j'aimerais faire une visite aux sites historiques de Geneve,166,3,1,1,228,1

je veux visiter les magasins de Genève,254,1,243,3,15,4

j'aimerais visiter Genève pour voir les boutiques de luxe,144,2,15,4,171,3

J'aimerais faire une visite familiale à Lausanne,17,4,227,2,24,3

je voudrais faire une visite culturelle à Genève,11,4,106,3,244,4

je cherche un restaurant à Lausanne,276,3,123,4,18,2

que puis je visiter à Genève avec ma famille,59,4,24,4,238,3

je voudrais visiter un musée d'art sur Genève,34,4,45,4,254,3

je veux me balader à pied à Genève,258,3,87,1,41,4

où est ce que je peux voir les paysages de Genève,18,4,213,1,246,1

je veux prendre le bateau a Geneve,18,1,1,1,238,2

visiter avec les enfants a geneve,38,4,78,2,92,1

j'aimerais me relaxer à Geneve,228,1,1,1,87,1

je peux voir quoi à Genève,238,2,246,1,213,1

je voudrais me promener dans les lieux culturelles de Genève,65,4,17,3,87,1

excursions à Geneve,224,3,1,1,231,3

Auriez vous quelques conseils de places à visiter dans un rayon raisonnable autour de Bern,238,4,184,2,160,1

je veux voir un musée typique suisse à Genève,73,1,216,1,45,3

Visite pour les curieux à Geneve,38,1,1,2,89,1

Les jolies paysages de Genève,37,4,143,3,69,1

Monuments de Genève,64,1,56,3,59,3

Visite religieuse,66,4,92,3,249,4

Spectacle à Genève,64,1,126,1,56,1

Belle vue sur Genève,59,4,137,3,66,4

Les adeptes de la science,36,4,62,4,127,1

Amoureux de la science,36,4,62,4,8,1

Visiter Geneve pour les amoureux de la nature,190,4,38,2,1,2

Symbole fort de Genève,31,4,130,3,97,2

Les amoureux de la litérature,36,2,190,2,124,2

Un chemin de promenade pour les passionnés de la nature,114,2,4,4,136,2

Un chemin de promenade pour les passionnés d'histoire,67,4,237,1,34,4

randonnées à Geneve,1,1,224,4,142,3

Une balade à travers des années d'histoire,238,2,246,4,272,4

Shopping Genève,64,4,56,1,59,1

Joli pont à Genève,56,2,40,3,111,3

Place très celebre à Genève,69,3,2,2,50,4

Amoureux de la littérature,67,3,54,3,117,2

Maisons célèbres,251,4,133,2,166,2

Jardin au centre de Genève,69,1,4,4,12,1

Panorama de Genève,64,1,56,4,59,4

Visite guidée Genève,47,2,19,4,66,4

Endroit historique au centre de Genève,69,4,56,4,154,3

Statue à Genève,58,4,64,1,56,2

Acheter du fromage suisse,73,4,82,4,260,2

déguster du fromage Suisse,73,4,210,4,76,3

Je veux manger au centre ville de Genève,70,4,12,2,77,4

Je veux manger traditionnel suisse,217,3,246,1,260,1

Je voudrais manger typique suisse à Genève,73,4,217,3,216,3

Je veux de la bouffe italienne à Genève,2,1,87,2,84,3

Recherche entrainé avec 50_50 questions:

je veux manger à Lausanne,160,3,121,4,113,2

j'aimerais faire une visite aux sites historiques de Geneve,166,3,1,1,38,1

je veux visiter les magasins de Genève,254,1,15,4,171,3

j'aimerais visiter Genève pour voir les boutiques de luxe,15,4,171,3,144,2

J'aimerais faire une visite familiale à Lausanne,17,4,24,3,227,2

je voudrais faire une visite culturelle à Genève,11,4,106,3,238,3

je cherche un restaurant à Lausanne,123,4,18,2,276,3

que puis je visiter à Genève avec ma famille,59,4,238,3,24,4

je voudrais visiter un musée d'art sur Genève,34,4,45,4,94,3

je veux me balader à pied à Genève,87,1,41,4,258,3

où est ce que je peux voir les paysages de Genève,18,4,254,1,213,1

je veux prendre le bateau a Geneve,18,1,238,2,17,4

visiter avec les enfants a geneve,78,2,38,4,92,1

j'aimerais me relaxer à Geneve,228,1,1,1,87,1

je peux voir quoi à Genève,238,2,246,1,254,2

je voudrais me promener dans les lieux culturelles de Genève,65,4,17,3,87,1

excursions à Geneve,224,3,1,1,231,3

Auriez vous quelques conseils de places à visiter dans un rayon raisonnable autour de Bern,238,4,184,2,46,2

je veux voir un musée typique suisse à Genève,73,1,45,3,7,4

Visite pour les curieux à Geneve,38,1,89,1,33,4

Les jolies paysages de Genève,69,1,203,3,143,3

Monuments de Genève,64,1,66,3,56,3

Visite religieuse,66,4,92,3,249,4

Spectacle à Genève,64,1,126,1,59,2

Belle vue sur Genève,59,4,137,3,66,4

Les adeptes de la science,36,4,62,4,127,1

Amoureux de la science,36,4,62,4,8,1

Visiter Geneve pour les amoureux de la nature,190,4,36,1,38,2

Symbole fort de Genève,31,4,130,3,97,2

Les amoureux de la litérature,190,2,36,2,124,2

Un chemin de promenade pour les passionnés de la nature,114,2,136,2,4,4

Un chemin de promenade pour les passionnés d'histoire,237,1,34,4,125,3

randonnées à Geneve,224,4,1,1,142,3

Une balade à travers des années d'histoire,238,2,246,4,272,4

Shopping Genève,64,4,59,1,66,2

Joli pont à Genève,56,2,40,3,167,3

Place très celebre à Genève,2,2,69,3,66,2

Amoureux de la littérature,67,3,54,3,117,2

Maisons célèbres,251,4,183,2,208,3

Jardin au centre de Genève,69,1,12,1,5,2

Panorama de Genève,64,1,66,4,56,4

Visite guidée Genève,66,4,19,4,47,2

Endroit historique au centre de Genève,69,4,56,4,154,3

Statue à Genève,58,4,64,1,56,2

Acheter du fromage suisse,73,4,82,4,226,4

déguster du fromage Suisse,73,4,226,4,76,3

Je veux manger au centre ville de Genève,70,4,79,4,77,4

Je veux manger traditionnel suisse,73,4,217,3,80,2

Je voudrais manger typique suisse à Genève,73,4,217,3,80,2

Je veux de la bouffe italienne à Genève,87,2,84,3,2,1

2. NOUVELLES QUESTIONS

Recherches classiques :

où est ce que je peux profiter du beau temps à Genève,246,1,213,1,243,2

visite pour les amateurs de l'art,244,4,176,4,34,4

visite en couple à Geneve, 1,3,71,2,146,3

Je veux manger une pizza à Geneve,79,3,1,1,74,4

Manger des glaces à Genève,75,4,199,3,38,1

Boire un verre dans un endroit charmant,245,4,250,3,241,4

Je désire manger asiatique,278,4,219,4,217,3

endroit romantique à Geneve,1,3,38,1,123,3

boutiques de luxe à Zurich,171,4,15,3,57,2

manger dans un fast-food à Zurich,160,3,46,1,217,4

faire du sky,238,2,226,4,232,3

faire du bateau à Zurich,17,4,238,4,219,3

Bronzer à Genève,56,1,64,1,59,2

se promener à Zurich,76,1,33,2,183,4

Chocolatier à Zurich,202,4,212,4,175,2

Chocolatier à Geneve,202,4,1,1,38,2

je veux manger des plats de la région à Zurich,276,3,35,1,147,1

je veux aller dans un bar à Zurich,218,4,243,2,242,2

je veux aller dans un bar à Genève,243,2,242,3,218,3

Faire des excursions en Valais,228,4,226,4,236,3

Excursions dans les Alpes,227,4,231,4,228,4

remontées mécaniques en Valais,230,4, 131,1,240,1

Vue des Alpes,35,4,229,4,181,2

je peux voir quoi dans la région Alpes,246,1,35,4,229,4

Les plus beaux chateaux de Suisse,17,1,10,1,233,1

belle vue sur Bale,256,2,59,3,181,3

les belles rues de Bale,256,2,183,3,153,3

Belles boutiques de Bale,256,2,265,3,270,4

Je veux voir les vieux quartiers de Bale,256,2,69,3,98,2

je veux boire un verre à Bale,256,1,245,4,18,3

boire un verre avec des amis,241,4,245,4,272,2

Visiter un musée à Zurich,255,3,182,2,45,3

Endroit à visiter à Zurich,166,4,184,182,4

Faire du sport à Zurich,189,4,107,2,125,1

Découvrir le chateau de Montreux,143,1,107,1,144,2

Je veux me promener à Montreux,228,2,87,1,17,4

Visite pour les musiciens,19,1,37,1,255,2

Boire un café à Lausanne,250,3,79,3,214,3

Découvrir Lausanne à pied,107,2,92,2,238,3

Repas asiatiques à Geneve, 278,3,1,2,38,2

Montreux pour les curieux,167,3,118,3,274,2— 274peut etre 1

contempler les paysages de Berne,155,4,193,1,153,2

lieu fort de Berne,155,4,230,2,156,4

Se promener à Zurich,76,2,33,3,183,4

Pont à Zurich,111,3,167,4,56,2

rivière de Zurich,172,4,155,3,217,2

Faire du shopping à Zurich,78,2,177,4,80,1

endroit pour les amateurs de la music,195,4,10,3,190,2

Endroit à ne pas manquer à Zurich,161,3,177,4,203,4

Recherche entrainé avec 50 questions:

où est ce que je peux profiter du beau temps à Genève,243,2,238,3,246,1

visite pour les amateurs de l'art,190,4,15,2,53,1

visite en couple à Geneve, 1,3,146,3,71,2

Je veux manger une pizza à Geneve,1,1,74,4,81,4

Manger des glaces à Genève,75,4,199,3,243,3

Boire un verre dans un endroit charmant,245,4,61,2,250,4

Je désire manger asiatique,219,4,217,3,79,3

endroit romantique à Geneve,123,4,1,3,136,2

boutiques de luxe à Zurich,144,2,171,4,15,3

manger dans un fast-food à Zurich,217,4,177,3,163,2

faire du sky,226,4,199,1,232,3

faire du bateau à Zurich,219,3,238,4,17,4

Bronzer à Genève,64,1,126,2,59,2

se promener à Zurich,183,4,65,3,243,3

Chocolatier à Zurich,212,4,185,2,184,1

Chocolatier à Geneve,1,1,126,1,212,3

je veux manger des plats de la région à Zurich,35,1,276,3,221,3

je veux aller dans un bar à Zurich,243,2,218,4,200,4

je veux aller dans un bar à Genève,243,2,218,3,123,3

Faire des excursions en Valais,224,4,226,4,231,4

Excursions dans les Alpes,35,4,223,4,202,1

remontées mécaniques en Valais,230,4,131,1,240,1

Vue des Alpes,35,4,233,4,140,4

je peux voir quoi dans la région Alpes,35,4,229,4,246,1

Les plus beaux chateaux de Suisse,17,1,10,1,233,1

belle vue sur Bale,59,3,263,3,181,3

les belles rues de Bale,258,3,203,3,256,3

Belles boutiques de Bale,256,2,177,3,171,3

Je veux voir les vieux quartiers de Bale,219,2,256,2,69,3

je veux boire un verre à Bale,256,2,18,3,250,4

boire un verre avec des amis,18,4,172,3,272,1

Visiter un musée à Zurich,255,3,34,3,185,2

Endroit à visiter à Zurich,184,4,186,4,171,4

Faire du sport à Zurich,189,4,107,2,184,1

Découvrir le chateau de Montreux,139,2,131,2,124,2

Je veux me promener à Montreux,17,4,228,2,87,1

Visite pour les musiciens,38,2,190,2,69,2

Boire un café à Lausanne,245,3,121,4,122,4

Découvrir Lausanne à pied,103,4,121,2,238,3

Repas asiatiques à Geneve,1,2,89,3,126,2

Montreux pour les curieux,235,3,144,2,167,3

contempler les paysages de Berne,140,3,143,3,18,3

lieu fort de Berne,155,4,156,4,97,3

Se promener à Zurich,76,2,183,4,115,3

Pont à Zurich,56,2,184,2,111,3

rivière de Zurich,217,2,172,4,203,3

Faire du shopping à Zurich,177,4,171,4,243,3

endroit pour les amateurs de la music,190,2,53,2,178,4

Endroit à ne pas manquer à Zurich,177,4,203,4,161,3

Recherche entrainé avec 75 questions:

où est ce que je peux profiter du beau temps à Genève,243,2,238,3,246,1

visite pour les amateurs de l'art,190,4,15,2,224,1

visite en couple à Geneve, 1,3,146,3,71,2

Je veux manger une pizza à Geneve,1,1,74,4,81,4

Manger des glaces à Genève,75,4,199,3,243,3

Boire un verre dans un endroit charmant,245,4,250,4,61,2

Je désire manger asiatique,217,3,79,3,219,4

endroit romantique à Geneve,123,4,1,3,136,2

boutiques de luxe à Zurich,171,4,144,2,15,3

manger dans un fast-food à Zurich,217,4,177,3,163,2

faire du sky,226,4,238,2,232,3

faire du bateau à Zurich,219,3,238,4,17,4

Bronzer à Genève,64,1,126,2,56,1

se promener à Zurich,183,4,65,3,243,3

Chocolatier à Zurich,212,4,185,2,166,2

Chocolatier à Geneve,1,1,38,2,89,2

je veux manger des plats de la région à Zurich,276,3,35,1,221,3

je veux aller dans un bar à Zurich,243,2,218,4,242,2

je veux aller dans un bar à Genève,243,2,218,3,242,3

Faire des excursions en Valais,224,4,231,4,226,4

Excursions dans les Alpes,223,4,35,4,202,1

remontées mécaniques en Valais,230,4,131,1,240,1

Vue des Alpes,35,4,140,4,233,4

je peux voir quoi dans la région Alpes,35,4,246,1,229,4

Les plus beaux chateaux de Suisse,17,1,233,1,10,1

belle vue sur Bale,59,3,263,3,181,3

les belles rues de Bale,187,1,256,3,203,3

Belles boutiques de Bale,256,2,171,3,15,3

Je veux voir les vieux quartiers de Bale,219,2,256,2,69,3

je veux boire un verre à Bale,18,3,250,4,256,2

boire un verre avec des amis,18,4,172,3,250,4

Visiter un musée à Zurich,45,3,34,3,23,3

Endroit à visiter à Zurich,184,4,186,4,171,4

Faire du sport à Zurich,189,4,107,2,175,1

Découvrir le chateau de Montreux,131,2,139,2,144,2

Je veux me promener à Montreux,17,4,228,2,87,1

Visite pour les musiciens,190,2,38,2,69,2

Boire un café à Lausanne,245,3,121,4,122,4

Découvrir Lausanne à pied,103,4,121,2,238,3

Repas asiatiques à Geneve,1,2,89,3,38,2

Montreux pour les curieux,235,3,144,2,167,3

contempler les paysages de Berne,140,3,37,3,143,3

lieu fort de Berne,156,4,155,4,97,3

Se promener à Zurich,76,2,183,4,115,3

Pont à Zurich,48,2,56,2,184,2

rivière de Zurich,172,4,217,2,203,3

Faire du shopping à Zurich,177,4,171,4,243,3

endroit pour les amateurs de la music,190,2,53,2,15,2

Endroit à ne pas manquer à Zurich,177,4,203,4,161,3

Recherche entrainé avec 100 questions:

où est ce que je peux profiter du beau temps à Genève,243,2,238,3,246,1

visite pour les amateurs de l'art,190,4,15,2,224,1

visite en couple à Geneve, 1,3,71,2,146,3

Je veux manger une pizza à Geneve,1,1,74,4,217,3

Manger des glaces à Genève,75,4,199,3,243,2

Boire un verre dans un endroit charmant,245,4,61,2,86,4

Je désire manger asiatique,278,4,219,4,217,3

endroit romantique à Geneve,123,4,1,3,136,2

boutiques de luxe à Zurich,171,4,144,2,15,3

manger dans un fast-food à Zurich,217,4,163,2,177,3

faire du sky,238,2,226,4,232,3

faire du bateau à Zurich,219,3,238,4,17,4

Bronzer à Genève,64,1,126,2,56,1

se promener à Zurich,183,4,65,3,243,3

Chocolatier à Zurich,212,4,185,2,166,2

Chocolatier à Geneve,1,1,212,3,126,1

je veux manger des plats de la région à Zurich,276,3,35,1,229,2

je veux aller dans un bar à Zurich,243,2,218,4,242,2

je veux aller dans un bar à Genève,243,2,242,3,218,3

Faire des excursions en Valais,224,4,231,4,226,4

Excursions dans les Alpes,35,4,223,4,202,1

remontées mécaniques en Valais,230,4,131,1,240,1

Vue des Alpes,35,4,229,4,140,4

je peux voir quoi dans la région Alpes,35,4,246,1,229,4

Les plus beaux chateaux de Suisse,17,1,233,1,10,1

belle vue sur Bale,59,3,263,3,256,3

les belles rues de Bale,256,3,203,3,258,4

Belles boutiques de Bale,256,2,15,3,171,3

Je veux voir les vieux quartiers de Bale,219,2,256,2,69,3

je veux boire un verre à Bale,256,2,18,3,250,4

boire un verre avec des amis,18,4,272,2,172,3

Visiter un musée à Zurich,45,3,34,3,255,3

Endroit à visiter à Zurich,184,4,171,4,186,4

Faire du sport à Zurich,189,4,107,2,184,2

Découvrir le chateau de Montreux,131,2,139,2,144,2

Je veux me promener à Montreux,228,2,17,4,87,1

Visite pour les musiciens,190,2,69,2,38,2

Boire un café à Lausanne,121,4,245,3,122,4

Découvrir Lausanne à pied,103,4,121,2,238,3

Repas asiatiques à Geneve,1,2,38,2,89,3

Montreux pour les curieux,167,3,235,3,118,3

contempler les paysages de Berne,140,3,37,3,143,3

lieu fort de Berne,156,4,155,4,97,3

Se promener à Zurich,76,2,183,4,115,3

Pont à Zurich,111,3,48,2,184,2

rivière de Zurich,172,4,217,2,203,3

Faire du shopping à Zurich,177,4,171,4,243,3

endroit pour les amateurs de la music,190,2,53,2,15,2

Endroit à ne pas manquer à Zurich,177,4,203,4,161,3

Recherche entrainé avec 50_50 questions:

où est ce que je peux profiter du beau temps à Genève,243,2,238,3,246,1

visite pour les amateurs de l'art,190,4,15,2,53,2

visite en couple à Geneve, 1,3,146,3,71,2

Je veux manger une pizza à Geneve,1,1,81,4,74,4

Manger des glaces à Genève,199,3,75,4,243,2

Boire un verre dans un endroit charmant,245,4,61,2,250,3

Je désire manger asiatique,217,3,79,3,80,4

endroit romantique à Geneve,123,4,1,3,136,2

boutiques de luxe à Zurich,171,4,144,2,15,3

manger dans un fast-food à Zurich,217,4,177,3,163,2

faire du sky,226,4,232,3238,2

faire du bateau à Zurich,219,3,17,4,238,4

Bronzer à Genève,64,1,126,2,59,2

se promener à Zurich,183,4,65,3,243,3

Chocolatier à Zurich,212,4,185,2,175,2

Chocolatier à Geneve,1,1,126,1,38,2

je veux manger des plats de la région à Zurich,35,1,276,3,121,3

je veux aller dans un bar à Zurich,218,4,243,2,200,4

je veux aller dans un bar à Genève,243,2,218,3,242,3

Faire des excursions en Valais,224,4,226,4,231,4

Excursions dans les Alpes,35,4,223,4,202,1

remontées mécaniques en Valais,131,1,240,1,230,4

Vue des Alpes,35,4,140,4,233,4

je peux voir quoi dans la région Alpes,35,4,229,4,246,1

Les plus beaux chateaux de Suisse,17,1,233,1,10,1

belle vue sur Bale,59,3,263,3,181,3

les belles rues de Bale,258,4,203,3,187,1

Belles boutiques de Bale,256,2,171,3,15,3

Je veux voir les vieux quartiers de Bale,219,2,69,3,256,2

je veux boire un verre à Bale,18,3,250,4,172,2

boire un verre avec des amis,18,4,272,2,172,3

Visiter un musée à Zurich,34,3,45,3,23,3

Endroit à visiter à Zurich,184, 186,4,4,171,4

Faire du sport à Zurich,107,2,189,4,175,1

Découvrir le chateau de Montreux,131,2,139,2,124,2

Je veux me promener à Montreux,17,4,228,2,87,1

Visite pour les musiciens,190,2,38,2,69,2

Boire un café à Lausanne,121,4,245,3,122,4

Découvrir Lausanne à pied,103,4,121,2,238,3

Repas asiatiques à Geneve,1,2,38,2,126,2

Montreux pour les curieux,235,3,167,3,118,3

contempler les paysages de Berne,37,3,143,3,140,3

lieu fort de Berne,156,4,155,4,97,3

Se promener à Zurich,76,2,183,4,115,3

Pont à Zurich,56,2,48,2,175,3

rivière de Zurich,172,4,217,2,203,3

Faire du shopping à Zurich,177,4,171,4,243,3

endroit pour les amateurs de la music,190,2,53,2,15,2

Endroit à ne pas manquer à Zurich,177,4,203,4,161,3